Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Fragment Binding to Kinase Hinge: If Charge Distribution and Local pK_a Shifts Mislead Popular Bioisosterism Concepts

*Matthias Oebbeke, Christof Siefker, Björn Wagner, Andreas Heine, and Gerhard Klebe**

anie_202011295_sm_miscellaneous_information.pdf

Author Contributions

MO and CS determined the crystal structures, BW measured the pK_a values in aqueous solution, AH and GK designed the study, MO and GK drafted the paper. All authors approved the final manuscript.

SUPPORTING INFORMATION

Table of Contents

Experimental Procedures

Protein expression and purification

The catalytic subunit of cAMP-dependent protein kinase (PKA) was derived from chinese hamster ovary (CHO) cells. A TEV protease cleavage site was introduced in a pET16b-vector between an His_7-tag and the CHO PKA N-terminus followed by a transformation of this plasmid into BL21 (DE3)/pLysS *E.coli* strain. CHO PKA was expressed in ZYM5052-medium.[1] After cell disruption with a high-pressure homogenizer the resulting cell lysate was centrifuged for 1h at 30.000 g. In a first step the supernatant was purified with a Ni-NTA column by applying a gradient of imidazole to 500 mM. After dialysis and cleavage of the His_7-tag with TEV protease an inverse Ni-NTA column was performed collecting the flow-through. As the final purification step a cation exchange column chromatography with MonoS™ (GE Healthcare) was performed and different phosphorylation states of CHO PKA were separated.

Crystallization and Soaking

For crystallization, the protein (0.2 mM) was transferred into a buffer containing 100 mM MES/Bis-Tris (pH 6.9), 75 mM LiCl, 0.03 mM mega 8, 1 mM DTT and 0.1 mM sodium EDTA. To perform vapour diffusion crystallization 3 µL of protein solution were pipetted on a stemp in sitting drop plates (Hampton Research). The reservoir was filled with 500 µL of a 23 % (v/v) methanol/water mixture. To initialize crystallization streak-seeding was performed by destroying old *apo* crystals with a horse hair and using these crystal fragments as seeds. Crystals could be observed after 3 weeks of growing and were transferred into a soaking buffer containing 10 % (v/v) of the fragment to be soaked (1 M in DMSO), 30 % (v/v) MPD and 60 % (v/v) crystallization solution. After 10-20 minutes crystals were flash frozen in liquid nitrogen and measured at the beamlines described in the section below (**Data Collection**).

Data Collection, Processing and Analysis

Crystals were measured at BESSY II on beamline 14.1 (**2**, **4** and **6**-**16**), 14.2 (**1**, **5** and **19**) and 14.3 (**3**).[2] Additional beamlines used for data collection were P11[3,4] at DESY (**17**) and ID23-2 at ESRF (**18**). Data processing was performed using XDS[5] and the molecular replacement was done with CCP4 Phaser[6] using a PDB structure of the PKA from *cricetulus griseus* (6F14). After a simulated annealing step in Phenix[7,8], refinement cycles (Phenix) and model building (Coot[9]) were carried out until convergence of free *R* factor (test set size of 5 % for all reflections) was reached. The structure's *B*-factors were refined anisotropically except those for the waters. In case of an at least 0.5 % higher free *R* factor value after anisotropic refinement, TLS refinement was chosen instead. Occupancies were refined for all atoms and multiple side-chain conformations were only build in, if the least occupied side chain amounted to 20 % or more. For structure validation, Ramachandran plots were calculated using PROCHECK[10] and temperature *B*-factors were calculated with Moleman[11,12]. All restraint files of the ligands were created by Mogul[13,14] on the Grade Web Server[15]. For structural analysis Pymol was used to prepare the figures.

Measurements of the pK_a values

The pK_a values were determined by SiriusT3 Fast UV pK_a method, which is based on the spectrophotometric (UV-metric) titration method reported in reference.[16] The compound of interest was prepared as 50 mM stock solution in DMSO and a fixed aliquot size of 3 µL was added to 1.5 mL of water containing 0.15 M KCl as background electrolyte. The pH of the dilute sample solution was adjusted to pH 2 by addition of 0.5 M HCl and then titrated with standardized base (0.5 M KOH) to pH 12 at 25°C under argon atmosphere. The SiriusT3 Fast UV pK_a method uses a proprietary linear buffer system adapted from the literature[17] to achieve much more rapid stabilization of the pH after each titrant addition. During the titration UV/vis spectra were collected as a function of the pH readings. The pK_a value of the sample is calculated from the pH readings and UV spectra collected.

Protein Data Bank Analysis with the GeoMine Tool

The tool GeoMine implemented in the Proteins.plus server from the Center of Bioinformatics at the University of Hamburg was used for the analysis of specific interaction types present in the PDB. Firstly nitro groups attached to an aromatic system were searched within the PDB, forming hydrogen bonds to a protein. Within the query a nitro group had to be defined using the ligand of 1D0C as a reference. Here the nitro group as well as the aromatic carbon atom next to this group were fixed in terms of bond lengths to define a nitro group bound to an aromatic system (SMARTS code: [$(a1aaaaa1)]), because this functional group was not implemented in this

tool. One of the two equivalent oxygens was defined as a hydrogen bond partner to a protein backbone or sidechain. The same was done for amidines defining an amidine group next to an aromatic system. For this query 1ZHR served as a reference structure.

Crystallographic Tables

All details about the datasets used in this publication are listed in table S1-S5.

Table S1. Crystallographic table for crystal structures obtained by soaking for ligands **1**-**4**.

Ligand (PDB code)	**1** (6SNX)	**2** (5N3H)	**3** (5N3Q)	**4** (5N3S)
Data Collection and Processing				
Number of crystals used	1	1	1	1
Wavelength (Å)	0.8655	0.9184	0.8950	0.9184
Space group	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters: a, b, c (Å)	51.8, 71.3, 97.2	51.3, 72.1, 97.4	51.6, 72.0, 97.6	51.7, 71.8, 97.4
Detector	DECTRIS PILATUS3S 2M	DECTRIS PILATUS3 6M	RAYONIX MX-225	DECTRIS PILATUS3 6M
Diffraction Data[a]				
Resolution range (Å)	45.67-1.40 (1.48-1.40)	50-1.36 (1.44-1.36)	50-1.31 (1.39-1.31)	50-1.14 (1.21-1.14)
Unique reflections	70483 (11148)	77685 (12245)	88063 (14042)	131934 (20912)
$R(I)_{sym}$ (%)	9.6 (45.0)	7.0 (60.6)	7.6 (53.2)	5.7 (56.6)
Completeness (%)	98.2 (97.7)	99.1 (98.0)	99.9 (99.5)	99.6 (98.6)
Redundancy	4.5	5.5	6.5	5.8
<I/σ(I)>	8.5 (2.2)	13.4 (2.4)	17.4 (5.1)	16.7 (3.5)
Refinement				
Resolution range (Å)	45.66-1.40	48.71-1.36	40.40-1.31	48.69-1.14
Reflections (work/free)	66855/3519	73775/3883	83658/4404	125337/6597
Final *R* value for all reflections (work/free) (%)	14.2/17.3	14.4/17.4	13.5/15.5	12.8/14.8
Protein	340	340	339	340
Inhibitor atoms	18	15	18/18	34
Water molecules	293	302	387	383
RMSD from ideality:				
Bond lengths (Å)	0.007	0.007	0.009	0.01
Bond angles (°)	0.9	0.9	1.1	1.1
Ramachandran plot (%)[b]				
Most favored regions	92.1	92.4	92.4	92.4
Additionally allowed regions	7.9	7.6	7.6	7.3
Generously allowed regions	0.0	0.0	0.0	0.3
Mean *B* factors (Å^2)[c]:				
Protein	18.0	18.0	13.6	14.4
Inhibitor	29.4	21.1	12.6/16.2	9.7
Water molecules	25.1	28.0	24.2	24.6
Matthews coefficient (Å^3/Da)[d]	2.2	2.2	2.2	2.2
Solvent content (%)[d]	43.7	45.0	45.0	45.0

[a] Values in parenthesis refer to the highest resolution shell
[b] Calculated using PROCHECK[10]
[c] Calculated using MOLEMAN[11,12] excluding hydrogen atoms
[d] Values calculated from Ruppweb[18]

SUPPORTING INFORMATION

Table S2. Crystallographic table for crystal structures obtained by soaking for ligands **5**-**8**.

Ligand (PDB code)	**5** (6Z44)	**6** (5N33)	**7** (6SPS)	**8** (6SPY)
Data Collection and Processing				
Number of crystals used	1	1	1	1
Wavelength (Å)	0.9184	0.9184	0.9184	0.9184
Space group	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters: a, b, c (Å)	51.8, 71.4, 97.4	58.2, 72.6, 108.8	51.7, 71.3, 97.5	51.5, 71.3, 97.5
Detector	DECTRIS PILATUS3S 2M	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M
Diffraction Data[a]				
Resolution range (Å)	48.71-1.38 (1.46-1.38)	50-1.43 (1.52-1.43)	45.65-1.65 (1.75-1.65)	48.75-1.60 (1.70-1.60)
Unique reflections	74984 (11843)	84316 (12943)	43348 (6900)	47330 (7445)
$R(I)_{sym}$ (%)	4.2 (49.7)	6.6 (52.6)	6.9 (49.5)	6.7 (48.4)
Completeness (%)	99.8 (98.9)	99.2 (95.3)	98.1 (97.9)	98.2 (97.1)
Redundancy	6.5	6.3	4.2	4.5
<I/σ(I)>	21.7 (3.1)	14.9 (2.6)	12.7 (2.3)	12.4 (2.4)
Refinement				
Resolution range (Å)	48.71-1.38	41.91-1.43	45.65-1.65	48.75-1.60
Reflections (work/free)	71233/3750	80100/4216	41179/2168	44963/2367
Final *R* value for all reflections (work/free) (%)	14.5/17.4	17.1/19.1	18.6/22.3	16.0/18.2
Protein/Peptide	338	351/13	338	338
Inhibitor atoms	7	17	13	15
Water molecules	258	323	254	308
RMSD from ideality:				
Bond lengths (Å)	0.007	0.008	0.006	0.008
Bond angles (°)	0.9	1	0.8	1.0
Ramachandran plot (%)[b]				
Most favored regions	91.7	92.8	90.4	91.1
Additionally allowed regions	7.9	7.2	9.3	8.9
Generously allowed regions	0.3	0.0	0.3	0.0
Mean *B* factors (Å^2) [c]:				
Protein/Peptide	19.6	24.5/29.2	24.0	21.8
Inhibitor	28.9	20.1	29.7	21.3
Water molecules	20.0	33.7	29.2	31.0
Matthews coefficient (Å^3/Da) [d]	2.2	2.8	2.2	2.2
Solvent content (%)[d]	44.0	57.0	43.8	43.7

[a] Values in parenthesis refer to the highest resolution shell
[b] Calculated using PROCHECK[10]
[c] Calculated using MOLEMAN[11,12] excluding hydrogen atoms
[d] Values calculated from Ruppweb[18]

SUPPORTING INFORMATION

Table S3. Crystallographic table for crystal structures obtained by soaking for ligands **9**-**12**.

Ligand (PDB code)	**9** (6SOX)	**10** (6SNN)	**11** (6SPM)	**12** (6SPU)
Data Collection and Processing				
Number of crystals used	1	1	1	1
Wavelength (Å)	0.9184	0.9184	0.9184	0.9184
Space group	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters: a, b, c (Å)	51.7, 71.4, 97.8	52.3, 71.5, 98.5	51.6, 71.4, 97.8	51.5, 71.6, 97.7
Detector	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M
Diffraction Data[a]				
Resolution range (Å)	45.70-1.38 (1.46-1.38)	46.17-1.82 (1.93-1.82)	45.61-1.37 (1.45-1.37)	45.55-1.39 (1.47-1.39)
Unique reflections	74638 (11910)	33571 (5195)	76337 (12112)	73002 (11715)
$R(I)_{sym}$ (%)	5.0 (48.2)	5.7 (50.0)	4.8 (48.5)	6.5 (48.1)
Completeness (%)	99.2 (99.2)	98.8 (95.9)	99.5 (99.0)	99.3 (99.5)
Redundancy	4.4	5.8	4.4	4.4
<I/σ(I)>	16.0 (2.7)	20.9 (3.2)	16.8 (2.7)	11.7 (2.3)
Refinement				
Resolution range (Å)	45.70-1.38	46.17-1.82	38.44-1.37	45.55-1.39
Reflections (work/free)	70906/3732	31892/1679	72520/3817	69351/3651
Final *R* value for all reflections (work/free) (%)	14.6/16.8	17.3/20.9	14.6/17.0	16.4/18.4
Protein	338	339	338	338
Inhibitor atoms	12	18	13	17
Water molecules	309	236	313	286
RMSD from ideality:				
Bond lengths (Å)	0.007	0.007	0.007	0.008
Bond angles (°)	1.0	0.8	0.9	1.0
Ramachandran plot (%)[b]				
Most favored regions	91.4	92.4	90.7	92.4
Additionally allowed regions	8.3	7.6	9.3	7.3
Generously allowed regions	0.3	0.0	0.0	0.3
Mean *B* factors (Å^2) [c]:				
Protein	16.0	24.3	16.0	18.9
Inhibitor	23.6	22.4	17.5	13.2
Water molecules	26.0	32.5	27.0	28.3
Matthews coefficient (Å^3/Da) [d]	2.2	2.2	2.2	2.2
Solvent content (%)[d]	44.1	45.2	44.0	44.0

[a] Values in parenthesis refer to the highest resolution shell
[b] Calculated using PROCHECK[10]
[c] Calculated using MOLEMAN[11,12] excluding hydrogen atoms
[d] Values calculated from Ruppweb[18]

SUPPORTING INFORMATION

Table S4. Crystallographic table for crystal structures obtained by soaking for ligands **13**-**16**.

Ligand (PDB code)	**13** (5N3J)	**14** (6Z08)	**15** (5N3E)	**16** (5N3D)
Data Collection and Processing				
Number of crystals used	1	1	1	1
Wavelength (Å)	0.9184	0.9184	0.9184	0.9184
Space group	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters: a, b, c (Å)	51.7, 71.5, 97.6	51.7, 71.3, 97.2	58.6, 72.3, 108.6	58.4, 72.4, 108.4
Detector	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M	DECTRIS PILATUS3 6M
Diffraction Data[a]				
Resolution range (Å)	50-1.12 (1.19-1.12)	45.61-1.49 (1.58-1.49)	50-1.53 (1.62-1.53)	50-1.77 (1.88-1.77)
Unique reflections	138256 (22096)	58203 (9369)	69062 (10849)	44971 (7147)
$R(I)_{sym}$ (%)	5.4 (57.2)	3.8 (50.2)	4.5 (64.7)	5.9 (53.4)
Completeness (%)	99.2 (99.0)	98.0 (99.0)	98.0 (96.3)	99.4 (99.0)
Redundancy	5.8	4.4	7.1	6.6
<I/σ(I)>	16.8 (3.0)	23.0 (2.9)	23.2 (3.5)	22.8 (3.7)
Refinement				
Resolution range (Å)	45.66-1.12	45.61-1.49	45.54-1.53	45.46-1.77
Reflections (work/free)	131343/6913	55292/2911	65608/3454	42722/2249
Final *R* value for all reflections (work/free) (%)	13.2/14.8	16.1/18.5	15.5/18.5	16.8/19.3
Protein/Peptide	340	338	352/14	350/13
Inhibitor atoms	16	10	21	21
Water molecules	367	276	241	223
RMSD from ideality:				
Bond lengths (Å)	0.010	0.008	0.006	0.007
Bond angles (°)	1.1	1.0	0.8	0.9
Ramachandran plot (%)[b]				
Most favored regions	91.8	92.4	92.6	93.4
Additionally allowed regions	8.2	7.6	7.1	6.6
Generously allowed regions	0.0	0.0	0.3	0.0
Mean *B* factors (Å^2) [c]:				
Protein/Peptide	14.3	20.6	31.7/34.0	28.7/32.2
Inhibitor	12.8	28.3	31.8	39.6
Water molecules	24.9	30.8	36.4	33.6
Matthews coefficient (Å^3/Da) [d]	2.2	2.2	2.8	2.8
Solvent content (%)[d]	45.0	43.6	57.0	56.0

[a] Values in parenthesis refer to the highest resolution shell
[b] Calculated using PROCHECK[10]
[c] Calculated using MOLEMAN[11,12] excluding hydrogen atoms
[d] Values calculated from Ruppweb[18]

SUPPORTING INFORMATION

Table S5. Crystallographic table for crystal structures obtained by soaking for ligands **17**-**19**.

Ligand (PDB code)	**17** (6YPS)	**18** (5N3C)	**19** (6ZN0)
Data Collection and Processing			
Number of crystals used	1	1	1
Wavelength (Å)	1.0332	0.8730	0.9184
Space group	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)	$P2_12_12_1$ (19)
Unit cell parameters: a, b, c (Å)	51.7, 71.3, 97.5	58.5, 72.7, 108.6	51.7, 71.2, 97.5
Detector	DECTRIS PILATUS 6M	MARMOSAIC 225 mm CCD	DECTRIS PILATUS3S 2M
Diffraction Data[a]			
Resolution range (Å)	48.75-1.35 (1.43-1.35)	50-1.77 (1.88-1.77)	45.70-1.59 (1.68-1.59)
Unique reflections	78185 (12184)	45623 (7178)	49029 (7675)
$R(I)_{sym}$ (%)	4.2 (52.9)	10.9 (59.5)	7.4 (49.0)
Completeness (%)	97.5 (95.1)	99.8 (98.9)	99.2 (97.3)
Redundancy	7.3	6.6	6.8
<I/σ(I)>	22.6 (3.5)	13.2 (3.1)	16.4 (4.2)
Refinement			
Resolution range (Å)	48.75-1.35	45.59-1.77	45.70-1.59
Reflections (work/free)	74275/3910	43341/2282	46577/2452
Final *R* value for all reflections (work/free) (%)	14.4/15.7	17.8/19.6	15.5/18.6
Protein/Peptide	338	352/13	338
Inhibitor atoms	10	15	9/9
Water molecules	236	235	273
RMSD from ideality:			
Bond lengths (Å)	0.008	0.007	0.005
Bond angles (°)	0.9	0.9	0.7
Ramachandran plot (%)[b]			
Most favored regions	91.4	92.2	91.7
Additionally allowed regions	8.6	7.8	8.3
Generously allowed regions	0.0	0.0	0.0
Mean *B* factors (Å^2) [c]:			
Protein/Peptide	20.7	25.8/31.3	17.8
Inhibitor	26.9	33.1	19.6/29.0
Water molecules	31.2	29.3	28.1
Matthews coefficient (Å^3/Da) [d]	2.2	2.8	2.2
Solvent content (%)[d]	43.9	57.0	43.8

[a] Values in parenthesis refer to the highest resolution shell
[b] Calculated using PROCHECK[10]
[c] Calculated using MOLEMAN[11,12] excluding hydrogen atoms
[d] Values calculated from Ruppweb[18]

SUPPORTING INFORMATION

Results

Crystal Structures

Before we describe all crystal structures in detail, a short summary of the water structures in the complexes should be given. Compared to the *apo* enzyme (5M0U) with three waters next to the hinge region, we mostly observed a displacement of these water molecules by the hinge binding ligands. In structures **3**, **4**, **6**, **12**, **15**, **16**, **17** and **19** a water molecule is located next to the amino acids Thr183 and Asp184 mediating interactions between protein and ligand. One exception is ligand **16** which is not interacting with this water molecule. In three structures (**6**, **11**, **19**) we see no displacement but slight shifts of a water molecule, which performs an interaction with Glu127 in the *apo* structure. Furthermore, ligand **18** does not directly bind to the hinge region, thus it is not displacing the two water molecules present in the *apo* structure. For all other crystal structures described in this study no conserved waters can be found.

We first crystallized benzamide **1** with PKA to complement to H-bonding pattern given by the hinge motif (Figure 1A). Benzamide interacts with the backbone carbonyl oxygen of Glu121 (2.9 Å) and the backbone nitrogen of Val123 (2.9 Å) of the hinge motif. The amide nitrogen is also able to interact with one of two conformers (47% populated) of the Thr183 side chain via a long H-bond (3.1 Å). Since the pK_a values of **1** are below 2 and above 12, (Table 1), an uncharged state of this ligand can be assumed under the applied ”soaking” conditions.

Nicotinamide **2** (or pyridine-3-carboxamide, Figure 1B) uses surprisingly its pyridine nitrogen to interact with the backbone nitrogen of Val123 (2.9 Å) as hydrogen acceptor (pK_a = 3.37). As second contact to the hinge, the backbone carbonyl oxygen of Glu121 is, water-mediated, involved in an H-bond to the amide nitrogen of **2**. In addition, the carbonyl oxygen connects the hinge region with the DFG motif of the kinase via 2.7 Å to Thr183Oγ. The well-defined electron density and the refined *B*-values show that the residual mobility of the fragment is much lower on the pyridine portion than on the carboxamide part of the ligand (pyridine-nitrogen: 15 Å²; carbonyl oxygen: 29 Å²; average: 21 Å²).

3-Aminobenzamide **3** (Figure 1C) interacts directly with the hinge motif with the expected pattern, similar to **1**. The carboxamide group points with two parallel H-bonds toward the backbone carbonyl oxygen of Glu121 and the backbone nitrogen of Val123. The distances are 2.9 Å for both H-bonds. The exocyclic amino group of **3** is at a distance of 3.1 Å to the hydroxyl group of Thr183 forming a third H-bond. The electron density of the fragment is clearly defined (15.6 σ (Fo-Fc)). The variation of *B*-values across **3** is low, from 10.0 Å² (benzene carbon next to the amide group) to 14.0 Å² (amino group).

4-Hydroxybenzamide **4** (Figure 1D) adopts two alternative orientation in the ATP-binding pocket (orientation A, amide group points to hinge; B, hydroxyl group points to hinge). In both orientations, **4** interacts with the hinge. Orientation A uses the carboxamide group, similar to **1** and **3**, to form two parallel H-bonds with the backbone nitrogen of Val123 and backbone carbonyl oxygen of Glu121. In orientation B, the phenolic hydroxy group, similar to the binding mode of phenol **5** (Figure 1E, s. below) serves simultaneously as H-bond donor and acceptor for the backbone carbonyl oxygen of Glu121 and backbone nitrogen of Val123. The hydroxyl group adopts twice the distance of 2.7 Å to both backbone atoms whereas the heteroatoms of the carboxamide group show two distances of 3.0 Å. In orientation B, the carboxamide group contacts the DFG loop with an H-bond (2.6 Å) via Thr183 and a second water-mediated interaction to the carboxylate oxygen of Glu170. It also interacts water-mediated with Glu127. This water molecule is at a distance of 3.1 Å to both groups. In contrast, the phenolic hydroxyl group in orientation A penetrates into the ribose pocket and interacts via a water bridge with the carboxylate group of Glu127 and with the backbone carbonyl oxygen of Leu49. The refined occupancy suggests orientation A to be populated by 60%, orientation B by 40%.

In contrast to a previously determined structure of phenol with human PKA, we observed only one phenol molecule **5** bound to CHO PKA (Figure 1E). It interacts with the hinge region in the above-described bifurcated fashion to Glu121C=O (2.7 Å) and Val123NH (2.8 Å). The pK_a value, reported in literature[19] suggests that under the applied "soaking" conditions, the equilibrium should favor uncharged species. This agrees with the observed interaction pattern, as phenol acts as donor and acceptor for hydrogen bonding.

Depending whether the fully aromatic tautomer or the one with an endocyclic amide bond is considered, fragment **6** comprises in the latter tautomer also a motif similar to a carboxamide moiety (Figure 1F). It does expose this motif to the hinge region and its amido N-H group is used to donate an H-bond to Glu121C=O (2.8 Å). The adjacent amide carbonyl oxygen accepts an H-bond from Val123NH (2.9 Å). The neighboring exocyclic amino group contacts a water molecule in the binding pocket. The CF_3 group, attached to the opposing face of the ring, orients toward the DFG loop. It is not in direct contact with these loop residues, but, mediated via a water molecule (3.1 Å), one of the fluorine atoms interacts with the backbone nitrogen atom of Asp184. The *B*-values suggest a higher residual mobility of the CF_3 group (22.8 Å²) compared to the remaining part of fragment **6** (average not considering the CF_3 group: 18.8 Å²).

The amide group of **7** interacts with the usual pattern (Figure 2A) to the hinge region via Glu121C=O (3.0 Å) and Val123NH (2.8 Å). The measured pK_a values suggests that **7** binds in uncharged state to the kinase. In addition, its *para*-trifluoromethyl group interacts through polar contacts with the sidechain of Glu127 (2.9 Å) and the backbone carbonyl oxygen (3.3 Å) and amide nitrogen (3.3 Å) of

Gly50. The CF_3 group shows high residual mobility compared to the remaining part of fragment **7** (compared *B*-values: carbon of CF_3 group 24.2 Å^2, carbon of carboxamide 12.9 Å^2).

With a pK_a value of 3.74, **8** binds most likely uncharged to the kinase (Figure 2B). It undergoes with its carboxamide group the usual interactions with the hinge (Glu121C=O, 3.1 Å and Val123NH 2.9 Å). The attached six-membered morpholino ring interacts weakly via its ether oxygen with a water molecule (3.4 Å).

Fragment **9** has, apart from the carboxamide group, in *para*-position a carboxylate function. For the interaction with the hinge, **9** uses its carboxamide establishing the usual interaction pattern (Figure 2C, Glu121C=O, 2.7 Å and Val123NH 2.9 Å). The orientation was mainly determined comparing the refined *B*-values for all oxygens and the nitrogen. Both oxygens of the carboxy group show nearly identical *B*-values, which indicates the fragment to be oriented correctly in the model (oxygens at carboxy group: 31.0 Å^2 and 31.0 Å^2, oxygen at carboxamide group 17.8 Å^2, nitrogen 19.1 Å^2). Compared to **8**, a shorter distance to the carbonyl group is experienced. Possibly, this correlates with the stronger electron withdrawing effect of the *para*-positioned acid group. The latter carboxylate group interacts weakly (3.5 Å) with the backbone carbonyl oxygen of Gly50 and water-mediated contact to Glu127 (2.5-2.8 Å) is found. With a measured pK_a value of 3.53 in aqueous solution, nevertheless the ligand could bind with protonated acid group to the protein as a result of a strong pK_a shift in the protein environment.

Considering the weakly acidic character of benzoic acids, binding in uncharged state to the protein cannot be excluded. Thus, we tested the binding of benzoic acid **10** (pK_a value = 4.01) and, remarkably, it interacts with identical interaction geometry as the carboxamide derivatives with the hinge (Figure 2D). This hydrogen-bonding pattern can only be realized if the carboxylic acid function serves either as hydrogen-bond donor and acceptor interacting with the carbonyl oxygen of Glu121 (2.6 Å) and the backbone nitrogen of Val123 (2.9 Å) in the usual pattern. Even though, the pK_a value of benzoic acid suggests full deprotonation in aqueous solution prior to protein bind, it must entrap a proton upon binding, which formally corresponds to a strong pK_a shift upon binding. Since in **9**, which features both functionalities carboxamide and carboxylic acid in the same molecule, the permanently uncharged carboxamide is preferred for the interaction with the hinge. Likely, for **10**, the effort to pick-up a proton upon binding has to be spend to accommodate the fragment next to the hinge.

The benzoic acid **11**, substituted in *para*-position with a CF_3 group, binds in the same way as the unsubstituted acid (Figure 2E). In addition, the acid group serves as donor (2.7 - 2.8 Å) and acceptor (2.9 Å) for the contact to the hinge. In addition, the electron-withdrawing trifluoromethyl group interacts with Glu127 (3.2 Å) and water-mediated with Glu170, similarly to the analogous carboxamide derivative **7**. Furthermore, there is a weak interaction with a water molecule at a distance of 3.4 Å. The attached CF_3 group shifts the pK_a value of the unbound fragment **11** to 3.48, possibly resulting in slightly shorter H-bonds to the hinge compared to the unsubstituted benzoic acid **10**. Similarly to **7**, the *B*-factor value of the central carbon of the carboxy function is much lower than the one of the CF_3 group (carbon of carboxylic acid: 14.9 Å^2, carbon of CF_3 group: 21.9 Å^2).

Adding an exocyclic amino group in *meta*-position to benzoic acid does not alter the binding pose (Figure 2F). The carboxylic acid function of **12** interacts as donor and acceptor with the hinge region (Glu121C=O: 2.6 Å, Val123NH 3.0 Å). The *meta*-amino group binds to the side chain of Thr183 (3.0 Å) and a water (3.0 Å). In aqueous solution, **12** exhibits pK_a values of 3.17 and 4.56, nevertheless, the interaction pattern suggests binding in uncharged state.

4-Nitrobenzoic acid (fragment **13,** pK_a = 3.20) adopts a rather surprising binding mode (Figure 3A). The carboxylate group of **13** should possess more acidic character due to the strong electron withdrawing effect of the *para*-NO_2 group. As a consequence, likely the fragment exposes only acceptor functionalities toward the protein. It might be due to this fact, that **13** binds with its nitro group to the hinge region. One oxygen atom of the nitro group interacts directly with the backbone nitrogen of Val123 (3.0 Å). The opposing carboxylic acid group interacts with three different protein residues. The terminal ammonium group of the side chain of Lys72 forms a salt bridge with **13** (2.8 Å). This underlines even more that the carboxylate group is present in deprotonated state. Mediated through a water molecule, the second carboxylate oxygen interacts with the backbone nitrogen of Phe185, which is part of the DFG motif. As third interaction partner, the backbone nitrogen of Asp184 forms an H-bond to the carboxylate group (2.9 Å, not shown). Additionally, the side chain of Met120 adopts an alternative orientation, which places the sidechain sulfur atom in 3.1 Å distance to the carboxylate oxygen. Fragment **13** exhibits consistent *B*-factor values for all atoms. To properly locate the acid and nitro groups of **13**, we compared the density peaks of the central atoms of both groups. The higher electron density peak allows placement of the nitro group. The crystal structure shows an exceptionally high resolution of 1.12 Å, which gives confidence to our assignment.

Interestingly, fragment **14** (nitrophenol, Figure 3B) avoids contacting the hinge, even though the hydroxyl function of the related phenol **5** establishes such an interaction. Instead **14** binds here with its hydroxyl function to the carboxylate group of Glu127 (2.9 Å). Obviously this ligand, which has a pK_a value of 6.91 in water, binds uncharged to the protein. The nitro group contacts the ammonium group of Lys72 via a weak hydrogen bond (3.2 Å).

Fragment **15** is again a benzoic acid, however now with a dimethyl amino substituent in *para*-position as electron pushing group (Figure 3C). Furthermore, its aromatic ring contains a pyridine-type nitrogen. The carboxylic acid must be present in protonated form and establishes the usual donor/acceptor hydrogen bonding pattern to Glu121C=O (2.6 Å) and Val123NH (2.8 Å). In addition, the pyridine-type nitrogen interacts water-mediated with Thr183Oγ and to one orientation (44% populated) of the side chain carboxylate group of Asp184. Via this contact, **15** connects the hinge region with the DFG loop. The tertiary amino group has no direct interaction partner and orients toward the opening of the binding pocket.

Fragment **16** adopts, overall, exactly the same binding pose as **15** (Figure 3D), even though **16** is a benzamidine and **15** a benzoic acid. This, on first glance astonishing bioisosteric replacement, is explained by the substituent patterns of both ligands. Both functional groups are present in uncharged state featuring a donor/acceptor pattern for the interaction with the hinge motif. One of the amidine nitrogen atoms interacts with the backbone nitrogen of Val123 as H-bond acceptor (2.9 Å). The second one acts as an H-bond donor with the backbone carbonyl oxygen of Glu121 (3.0 Å). Usually, amidines have strong basic character with pK_a values well beyond 12. The CF_3 group however, reduces the basic character via its -I effect, resulting in binding of **16** in uncharged state. Accordingly, a pK_a value of 10.78 is measured in water for **16**. Two of the fluorine atoms are in close contact with the protein and show quite similarly geometries as the other CF_3 containing fragments **7** and **11**. Comparing the *B*-factor values of these two fragments with **16** shows the same trend that the CF_3 group exhibits much higher values than the group binding to the hinge (carbon of CF_3 group: 48.6 Å^2, carbon of carboximidamide group: 28.3 Å^2).

Fragment **17** binds with its amidino group to the hinge (Figure 3E), however only one nitrogen is used to contact Val123NH at a distance of 3.0 Å. If the expanded distance displays a proper H-bond, the amidino function must be present in uncharged state. Only then, the group can act as hydrogen-bond acceptor, despite a pK_a value beyond 12 of this fragment in aqueous solution. It is surprising that a deviating binding mode is adopted compared to **16**, even though both fragments seem to exhibit the same pronation state. The additional *para*-hydroxyl group of **17** (pK_a value = 7.75) interacts via a water (2.5 Å) with Lys72 (3.1 Å).

Fragment **18** (Figure 3F) exhibits an amidino group with a basic pK_a value of 11.32 in aqueous solution. Likely, **18** binds in charged state to the protein and avoids contacting the hinge region. The protonation of the amidino group is supported by the observation that **18** forms a salt bridge with the side chain carboxylate group of Asp184 (part of the DFG motif) and H-bond distances of 2.8 and 2.7 Å are observed. Fragment **18** is to 69 % populated in the binding pocket.

Finally, we studied *iso*-nicotinamidine **19** (Figure 3G). Also this fragment comprises an amidino group with a pK_a value of 10.12. Remarkably, it also does not expose the latter functional group toward the hinge as seen for **16** and **17**. Instead, similarly to **2** (Figure 1B), it uses the endocyclic pyridine-type nitrogen (pK_a value 2.23) to form an H-bond to Val123NH of the hinge with 3.0 Å distance. The opposing *para*-amidino group donates an H-bond to Thr183Oγ (2.9 Å). With this, it differs from **18**, which also orients the amidino group in this area but selects Asp184 as interaction partner. In contrast to **2**, fragment **19** is spatially rotated and forms two H-bonds to water molecules, which are located at different spots compared to the complex with **2**. The protonation state of the amidino group is difficult to estimate. However, as three H-bonds of rather long distance are experienced, the uncharged state appears likely. Similar to fragment **2**, the refined *B*-values indicate lower residual mobility of the fragment next to the pyridine portion (e.g. pyridine-nitrogen: 17.4 Å^2; amidino nitrogens: 22.6 and 23.6 Å^2; average for all atoms of **2**: 19.6 Å^2).

An analysis of all structures in the PDB by the GeoMine tool (description see above in the experimental section) was done to obtain an overview of the interactions found in this contribution for a nitro group attached to an aromatic ring system compared to the occurrence of such an interaction present in general in the PDB. 506 matches were found in 399 pockets of 214 PDB entries. On first glance, among these matches ligands might serve in four cases as H-bond donor and in 501 cases as an acceptor for hydrogen bonding. From a sole visual inspection of these H-bond donors, it remains unclear whether they really serve as donors, also whether possibly the corresponding functional group on the side of the protein has changed protonation state. A refinement of this search resulted in two structures (2W7X, and **13** from this paper 5N3J) where ligands form hydrogen bonds with their nitro group to the hinge of a kinase. For a similar analysis of amidines attached to an aromatic system, the search resulted in 513 interactions present in 282 pockets of 246 PDB entries. Both amidine nitrogens were considered to be identical and served as H-bond donors in all cases. As in the nitro example above, we have to be careful about the obtained interaction statistics, because the algorithm could list ligands in, e.g., dimeric proteins which would count twice summing up redundant information. It could therefore be more conclusive to compare the numbers of PDB entries found to contain the specific interaction. Furthermore, it is questionable whether the GeoMine algorithm can correctly predict the functional group of a ligand to serve either as H-bond donor or acceptor. A similar problem might exist for the assignment of protonation states to residues such as His, Asp, Glu, Tyr, or Lys. Nevertheless the interactions found with this tool show that the two types of interactions, described above, can be found in many structures deposited in the PDB.

Abbreviations

Bis-Tris: 2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol, CHO: chinese hamster ovary, DMSO: dimethyl sulfoxide, DTT: dithiothreitol, EDTA: Ethylenediamintetraacetic acid, His_7-tag: histidine-tag, MES: 2-(*N*-morpholino)ethanesulfonic acid, MPD: 2-

methyl-2,4-pentanediol, Ni-NTA: nickel-nitrilotriacetic acid, PDB: protein data bank, PKA: cAMP-dependent protein kinase A, TEV: tobacco etch virus.

References

[1] F. W. Studier, *Protein expression and purification* **2005**, *41*, 207.

[2] Helmholtz-Zentrum Berlin für Materialien und Energie. The MX beamlines BL14.1-3 at BESSY II, *Journal of large-scale research facilities* **2016**, *2*, A47.

[3] A. Burkhardt, T. Pakendorf, B. Reime, J. Meyer, P. Fischer, N. Stübe, S. Panneerselvam, O. Lorbeer, K. Stachnik, M. Warmer et al., *Eur. Phys. J. Plus* **2016**, *131*.

[4] A. Meents, B. Reime, N. Stuebe, P. Fischer, M. Warmer, D. Goeries, J. Roever, J. Meyer, J. Fischer, A. Burkhardt et al. in *SPIE Proceedings* (Ed.: B. Lai), SPIE, **2013**, 88510K.

[5] W. Kabsch, *Acta crystallographica. Section D, Biological crystallography* **2010**, *66*, 125.

[6] 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.

[7] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy et al., *Acta crystallographica. Section D, Structural biology* **2019**, *75*, 861.

[8] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta crystallographica. Section D, Biological crystallography* **2012**, *68*, 352.

[9] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta crystallographica. Section D, Biological crystallography* **2010**, *66*, 486.

[10] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J Appl Crystallogr* **1993**, *26*, 283.

[11] G. J. Kleywegt, "MOLEMAN. Unpublished Program; Uppsala University: Uppsala, Sweden".

[12] G. J. Kleywegt, J.-Y. Zou, M. Kjeldgaard, T. A. Jones in *International Tables for Crystallography* (Ed.: H. Fuess), Springer Netherlands, Amsterdam, **2006**, pp. 353–356.

[13] I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris et al., *Journal of chemical information and computer sciences* **2004**, *44*, 2133.

[14] S. J. Cottrell, T. S. G. Olsson, R. Taylor, J. C. Cole, J. W. Liebeschuetz, *Journal of chemical information and modeling* **2012**, *52*, 956.

[15] O. S. Smart, T.O. Womack, A. Sharff, C. Flensburg, P. Keller, W. Paciorek, C. Vonrhein, G. Bricogne, Cambridge, United Kingdom, Global Phasing Ltd., "Grade Web Server", can be found under http://www.globalphasing.com, **2011**.

[16] K. Y. Tam, K. Takács-Novák, *Analytica Chimica Acta* **2001**, *434*, 157.

[17] K. Box, C. Bevan, J. Comer, A. Hill, R. Allen, D. Reynolds, *Analytical chemistry* **2003**, *75*, 883.

[18] Bernhard Rupp, "Matthews Probability Calculator", can be found under http://www.ruppweb.org/mattprob/default.html, **2018**.

[19] D. R. Lide (Ed.) *CRC handbook of chemistry and physics*, CRC Press, Boca Raton, **2004**.

Author Contributions

MO and CS determined the crystal structures, BW measured the pK_a values in aqueous solution, AH and GK designed the study, MO and GK drafted the paper. All authors approved the final manuscript.