Supporting Information

Virtual Screen to NMR (VS2NMR): Discovery of fragment hits for the CBP bromodomain

Dimitrios Spiliotopoulos,[*,a] Jian Zhu,[a] Eike-Christian Wamhoff,[b,c] Nicholas Deerain,[a] Jean-Rémy Marchand,[a] Jonas Aretz,[b,c] Christoph Rademacher,[b,c] and Amedeo Caflisch[*,a]

Author Affiliations:
[a]Department of Biochemistry University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
[b]Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14424 Potsdam, Germany
[c]Institute of Chemistry and Biochemistry, Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Takustraße 3, 14195 Berlin, Germany

Corresponding Authors:
*E-mail: A.C., caflisch@bioc.uzh.ch. Phone: +41 44 635 55 21. D.S., d.spiliotopoulos@bioc.uzh.ch. Phone: +41 44 635 55 92.

EXPERIMENTAL SECTION

Fragment Docking and Scoring. The initial library consisted of 1413 fragments available in the laboratory of one of the authors (C.R.). These molecules were selected from a large panel of commercial suppliers and academic collaborations according mainly to diversity. From this library, a total of 2133 tautomers were generated using the calculator plugins of Marvin 15.8.17, 2015, Chemaxon.

The in-house developed program SEED[1, 2] was used for docking. The target structure was the CBP bromodomain complexed to acetylated lysine (PDB code: 3P1C), and the binding site was defined as the side chain of the conserved Asn1168 and the six water molecules that are found in most crystal structures of this bromodomain (Figure S1).

The partial charges and van der Waals parameters for the protein and the fragments were taken from the CHARMM36 all-atom force field[3, 4] and the CHARMM general force field (CGenFF),[5] respectively. Importantly, the same paradigm was used to derive the partial charges and van der Waals parameters for the CHARMM36 force field and CGenFF, making the force fields completely consistent. SEED uses a force field-based energy function with a continuum dielectric approximation of desolvation penalties based on the generalized Born paradigm to evaluate the binding energy.[6] The continuum calculations were performed setting the dielectric constant to 2.0 and 78.5 for the solute (low-dielectric region) and solvent (high-dielectric region), respectively. The docking of the 2133 tautomers with SEED required approximately 2 h (about 3 s per fragment) of a single core of an Intel Xeon E5410 processor at 2.33 GHz.

The docked poses were first evaluated for the presence of an acceptor atom involved in a hydrogen bond with the conserved water molecule w_1 (Figure S1). Moreover, poses with buried polar groups of fragment and/or protein not involved in hydrogen bonds were filtered out using an in-house developed software (Hydrogen bond penalty lower than 1).[7, 8] The final ranking was based on the median value of a consensus scoring function that included the ranks of (1) the difference between (1a) the electrostatic contribution to the protein/ligand interaction energy in the solvent and (1b) the solvation energy of the ligand, (2) the predicted binding energy (SEED total energy), and (3) the van der Waals efficiency, i.e., the intermolecular van der Waals contribution divided by the number of non-hydrogen atoms. These three terms are meant to prioritize compounds that (1) establish favorable polar interactions with the targeted protein considering the opposing free energy of hydration, (2) have favorable total energy calculated by SEED, including van der Waals and polar interactions, and (3) are fully buried in the binding site. Compounds were sorted using the median of the three rankings. The median (and not the mean value) was used as it is less sensitive to outliers.[9, 10] Note that SEED treats the docked compounds as rigid molecules, meaning that all the terms (including 1a and 1b) are computed for the small molecule in the conformation under investigation. A total of 60 small molecules were selected using the *in silico* approach. Of these 60 compounds, 21 were subsequently filtered out due to experimental issues (e.g., poor solubility, binding promiscuity and/or chemical reactivity).

Protein Expression and Purification. Proteins were expressed and purified as described in [11].

NMR Spectroscopy. All NMR experiments were performed on a PremiumCompact 600 MHz spectrometer at 25°C equipped with a OneNMR probe (Agilent). Data were processed using MestReNova software suite (Mestrelab Research S. L.). A DPFGSE pulse sequence was utilized for solvent suppression.[12]

During STD NMR experiments a saturation time t_{sat} of 4.0 s utilizing a train of 50 ms Gauss pulses for used with an on- and off-resonance frequency of 0.0 ppm (v_{sat}) and 80.0 ppm (v_{ref}).[13] For each spectrum, 256 scans were recorded in 5 mm sample tubes at sample volumes of 500 to 550 µL. No prescan relaxation delay d_1 was included and the acquisition time t_{acq} was set to 2.0 s. A $T_{1,rho}$ filter of 35 ms duration was utilized for receptor resonances suppression. A Carr−Purcell−Meiboom−Gill (CPMG) pulse sequence was used to perform R_2-filtered NMR experiments with a prescan relaxation delay d_1 of 2.0 s and an acquisition time t_{acq} of 2.0 s.[14] The frequency of 180° pulses (v_{CPMG}) was set to 100 Hz and the total relaxation time T was set to 0.4 s. The induction of chemical shift perturbations in presence of CBP was analyzed using regular ^{1}H NMR experiments. The relaxation delay d_1 was set to 2.0 s and the acquisition time t_{acq} was set to 2.0 s. Spectra were recorded at 128 scans.

From the 39 selected compounds, four fragment mixtures with minimal 1H NMR spectral overlap were predicted using a genetic algorithm. Briefly, this approach was implementing by first generating a list of chemical shifts for each fragment. Assignment of the NMR resonances was based on previously acquired ^{1}H NMR spectra of the individual fragments. The lists were then randomly combined to yield a population of 50 individual virtual fragment mixtures. Assuming a tolerance of 0.3 ppm for the spectra overlap, a fitness score was calculated. Based on this score, a set of fit fragment mixtures is selected and its composition is randomly varied by mutation (mutation rate = 0.01) or crossover (crossover rate = 0.8) to generate the next generation of fragment mixtures. This procedure was repeated for 1000 generations until convergence to yield fragment mixtures with optimized spectral overlap. The termination criterion was the existence of at least one resonance not displaying spectral overlap for each fragment.

Optimized sample mixtures were prepared at 200 µM of each fragment in 50 mM H_3PO_4 with 100% D_2O, 2% DMSO-d_6, 150 mM NaCl at pH 7.4. 100 µM TSP-d_4 served as an internal reference. The stability of the fragment mixtures over 16 h at room temperature was monitored via ^{1}H NMR. CPMG and STD-based fragment screening was conducted first in absence of CBP, followed by the addition of 20 µM CBP. Finally, competitive binding experiments were conducted in presence of 20 µM SGC-CBP30.[15] Hits from virtual screening were considered as validated in case either an STD effect, an increased R_2 relaxation rate or a ^{1}H chemical shift perturbation occurred in presence of CBP, and the nanomolar inhibitor SGC-CBP30 showed competitive binding. The analysis was based on a visual and qualitative assessment of the recorded spectra. Fragments displaying an STD effect in absence of protein were considered artifacts and were excluded from further analysis.

AlphaScreen and BROMOscan assays. AlphaScreen is a bead-based proximity assay technology that has been applied to identify small molecules able to displace histone peptides from bromodomains.[16] Briefly, donor and acceptor beads are coupled to the interaction partners and, as the bromodomain ligand inhibits this interaction, the detected signal is reduced. A histone H4 tetra-acetylated peptide (H4$_{1-21}$Kac5Kac8Kac12Kac16) was used for the measurements with the CBP bromodomain ligands in the presence of 0.1% DMSO.
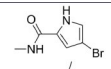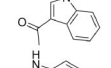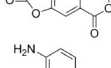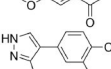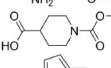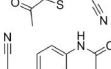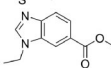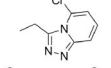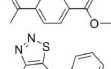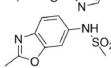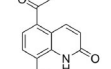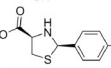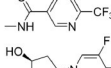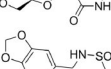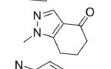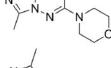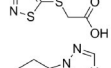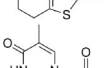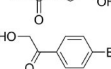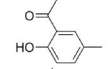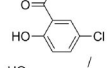
BROMOscan is a competition-based technology using a ligand immobilized to a solid support and DNA-tagged bromodomains. The ligand is incubated with the bromodomains in the presence and absence of the putative inhibitors and the bromodomains are eluted and quantified by qPCR. The amount of bromodomain captured will be reduced if small molecules inhibiting the bromodomain binding to the immobilized ligand are present, which results in a reduction of the detected qPCR signal.[17] Dissociation constants (K_D) were calculated in the presence of 0.09% DMSO fitting a 12-point dilution with starting concentration of 0.5 mM and dilution factor of 3.0.
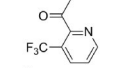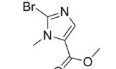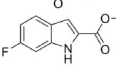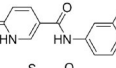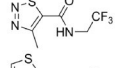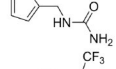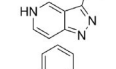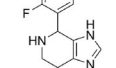
AlphaScreen and BROMOscan assays were carried out at Reaction Biology Corporation and DiscoverX Corporation, respectively.

Thermal Shift Measurements. Thermal shift measurements were performed using a 2 µM and 1 mM concentration for the bromodomains and ligands, respectively, with 0.3% (v/v) DMSO.

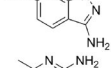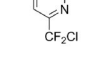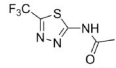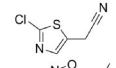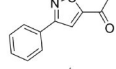**Crystallization, Data Collection and Structure Solution**. CBP bromodomain was crystallized by vapor diffusion in sitting drops at 4°C. Apo crystals for compound soaking were grown from 0.1 M Bis-Tris pH 5.5, 0.2 M potassium thiocynate, 5% ethylene glycol and 23% PEG3350. Overnight soaking of compounds **1** and **4** were performed by transferring the apo crystals into the crystallization buffer in which compounds were previously dissolved at 5-10 mM. Compound **2** was co-crystallized with CBP bromodomain using the crystallization buffer of 0.15 M KSCN, 10% ethylene glycol and 20% PEG3350. Co-crystal structure of compound **3** bound to CBP bromodomain was determined from crystals grown from 0.1 M Bis-Tris pH 6.5, 0.2 M MgCl$_2$, 5% ethylene glycol and 23% PEG3350. Crystals were cryoprotected by crystallization buffer supplemented with 20% ethylene glycol prior to freezing in liquid nitrogen. Diffraction data from a single crystal was acquired at the X06SA beamline at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. Data was processed with XDS[18] and scaled with SCALA[19] or AIMLESS,[20] structures were solved by molecular replacement with Phaser[21] or MOLREP[22] using PDB 3DWY as search model. Clear difference electron densities for compounds in the Kac binding pocket can be unambiguously modelled. Rounds of manual model building were carried out with COOT[23] and refinement was performed with Phenix.[24] Crystal data collection and refinement statistics are summarized in Table S2.

Table S1. 2D structures and contributions to the binding energy (in kcal/mol) for the 39 molecules predicted as candidate ligands of the CBP bromodomain by the docking program SEED[a].

	2D structure	intermolecular		electrostatic desolvation		total	Total
		vdW	elect.	protein	fragment	elect.	energy
1		−16.9	−11.0	5.1	6.0	0.1	−16.8
2		−17.9	−7.7	6.7	3.4	2.4	−15.5
3		−14.5	−9.4	6.1	2.6	−0.7	−15.2
4		−18.0	−10.3	7.2	3.3	0.2	−17.8
5		−18.2	−15.2	8.2	8.7	1.7	−16.5
6		−20.1	−9.7	6.9	6.0	3.2	−16.9
7		−16.1	−7.7	6.9	3.0	2.2	−13.9
8		−14.8	−10.9	6.4	4.4	−0.1	−14.9
9		−19.1	−10.2	8.0	2.9	0.7	−18.4
10		−22.6	−7.5	5.6	4.8	2.9	−19.7
11		−16.8	−9.4	6.5	2.8	−0.1	−16.9
12		−22.2	−5.9	7.8	2.2	4.1	−18.1
13		−18.6	−5.0	7.5	2.7	5.2	−13.4
14		−20.7	−12.2	8.3	6.6	2.7	−18.0
15		−17.8	−21.9	11.2	15.1	4.4	−13.4
16		−17.1	−11.5	5.6	5.2	−0.7	−17.8
17		−19.2	−12.4	8.9	9.1	5.6	−13.6
18		−15.3	−3.0	6.5	4.0	7.5	−7.8
19		−17.5	−8.1	5.3	2.6	−0.2	−17.7
20		−19.5	−10.1	9.0	4.4	3.3	−16.2
21		−20.7	−9.4	6.2	5.3	2.1	−18.6
22		−20.8	−8.3	6.1	1.6	−0.5	−21.4
23		−17.8	−11.1	6.7	8.2	3.8	−14.0
24		−16.3	−12.1	5.6	5.0	−1.5	−17.8
25		−18.1	−12.5	6.5	6.2	0.2	−17.9
26		−18.4	−12.4	6.4	6.3	0.3	−18.0
27		−17.9	−8.7	5.8	7.5	4.6	−13.3
28		−18.4	−6.2	6.0	5.9	5.7	−12.7

29		−16.2	−12.8	5.2	8.9	1.3	−14.9
30		−17.7	−6.3	5.2	2.9	1.8	−15.9
31		−16.3	−8.0	6.0	2.8	0.8	−15.5
32		−15.5	−7.4	5.4	2.8	0.8	−14.7
33		−18.3	−7.0	5.7	2.8	1.5	−16.8
34		−16.5	−9.7	6.0	3.4	−0.3	−16.8
35		−17.3	−16.2	8.9	7.4	0.1	−17.2
36		−17.7	−7.3	7.4	2.4	2.5	−15.2
37		−11.9	−14.4	5.6	5.6	−3.2	−15.1
38		−11.0	−9.0	4.9	3.5	−0.6	−11.6
39		−15.9	−14.3	9.1	6.3	1.1	−14.8

[a]The SEED total energy (total energy) is the sum of the intermolecular van der Waals energy (vdW) and the total electrostatic energy. The total electrostatic is the sum of intermolecular electrostatic energy and desolvation penalties for protein and fragment.

Table S2. X-ray data collection and refinement statistics for complex structures of the CBP bromodomain and compounds 1, 2, 3 and 4.

Data Collection				
PDB ID	5MQE	5MQK	5MPZ	5MQG
ligand	**1**	**2**	**3**	**4**
space group	H3	H3	$P2_12_12_1$	H3
Cell dimensions				
a, b, c (Å)	122.37, 122.37, 39.35	122.36,122.36, 40.37	44.06, 44.06, 60.21	121.44, 121.44, 40.27
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 90.00	90.00, 90.00, 120.00
resolution (Å)	37.38 - 1.65	37.72 - 1.53	44.06 - 1.40	37.61 - 1.35
unique observations[a]	26769(1347)	33860(4964)	23766(1057)	48059(6756)
completeness[a]	99.7(99.5)	99.6(99.9)	98.9(90.1)	98.8(95.4)
redundancy[a]	5.1(4.5)	5.1(4.9)	5.8(3.3)	4.9(4.2)
Rmerge[a]	0.038(0.425)	0.041(0.485)	0.044(0.336)	0.030(0.335)
I/σI[a]	20.2(3.0)	16.5(3.3)	21.2(4.0)	20.7(3.6)
Refinement				
Rwork/Rfree[a]	0.184(0.260) /0.223(0.322)	0.183(0.278)/ 0.210(0.275)	0.157(0.206)/ 0.176(0.273)	0.171(0.244) /0.191(0.252)
r.m.s. deviations of bond lengths (Å)	0.007	0.007	0.005	0.006
r.m.s. deviations of bond angles (°)	0.909	0.915	0.957	0.891
Average B-factor (Å^2)				
protein	36.13	37.04	17.66	28.87
ligand	24.69	30.31	24.05	27.71
water	40.83	41.61	31.76	37.04
Ramachandran				
favored (%)	98.57	99.06	100	99.50
allowed (%)	1.43	0.94	0	0.50
disallowed (%)	0	0	0	0

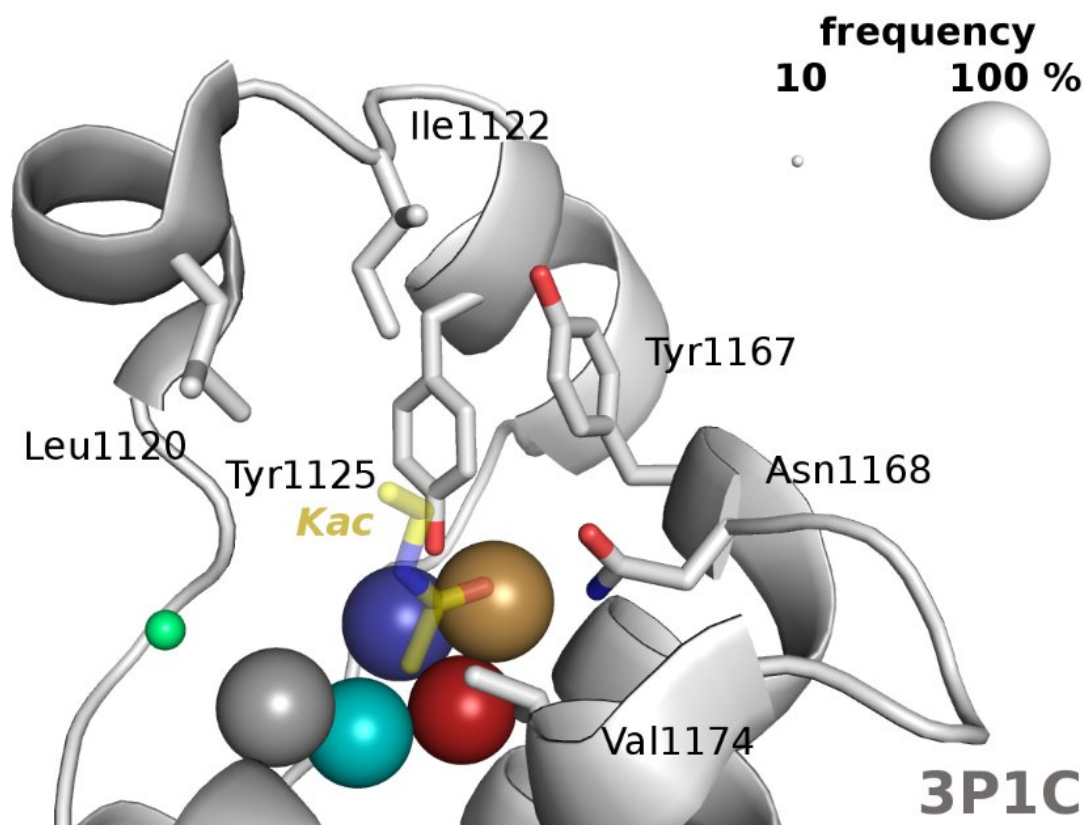[a]Highest resolution shell is shown in parentheses.

Figure S1. CBP bromodomain structure used for docking (PDB code: 3P1C). The six water molecules kept for the docking procedures are shown as spheres, with w_1 colored in gold (color code for the water molecules as in Figure 1B of [25]). The size of each sphere is proportional to the frequency of the corresponding water molecule (legend on top, right) in 44 inspected crystal structures of bromodomains bound to an acetyllysine (one structure each for PCAF, BRD2(2), BRD3(2), BRD4(2), BRDT(1), BRDT(2), and BAZ2A; two structures for TAF1(2); three structures each for CBP, BAZ2B, and TRIM24; four structures each for BRD2(1), ATAD2, and BRPF1; seven structures each for BRD4(1) and BRD9). The side chain of the residues discussed in the main text and the acetyllysine are shown as sticks (carbon atoms in gray and yellow, respectively).

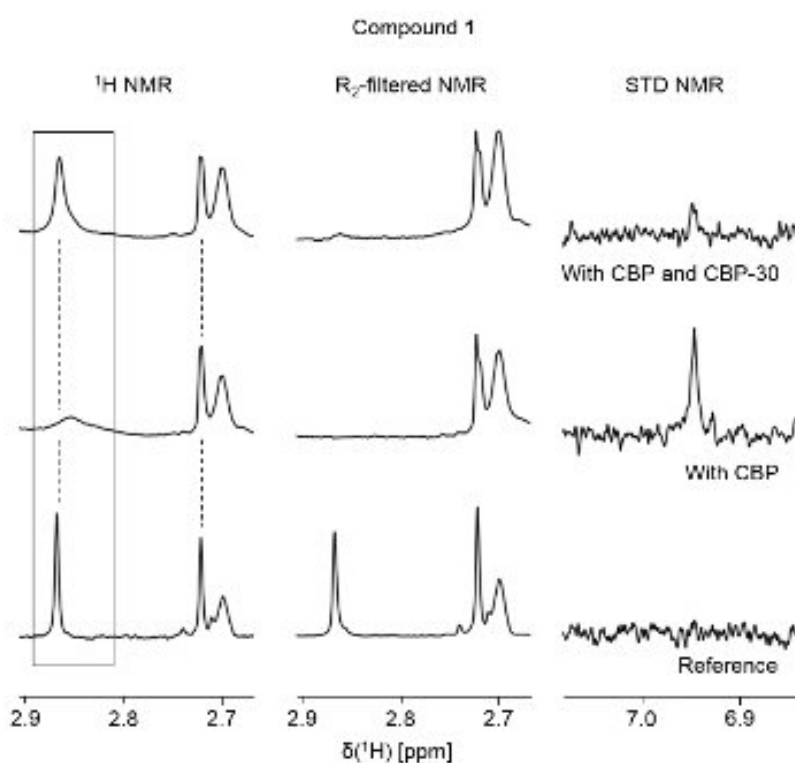**Figure S2**. Ligand-observed NMR screening experiments for compound **1**. Peaks of the compound undergo ^{1}H chemical shift perturbation (left), an increase in the R_2 relaxation (middle) and STD effects (right) in the presence of the CBP bromodomain. Binding occurs in the acetyllysine binding site as titration of the nanomolar inhibitor SGC-CBP30 reverses all three observables.
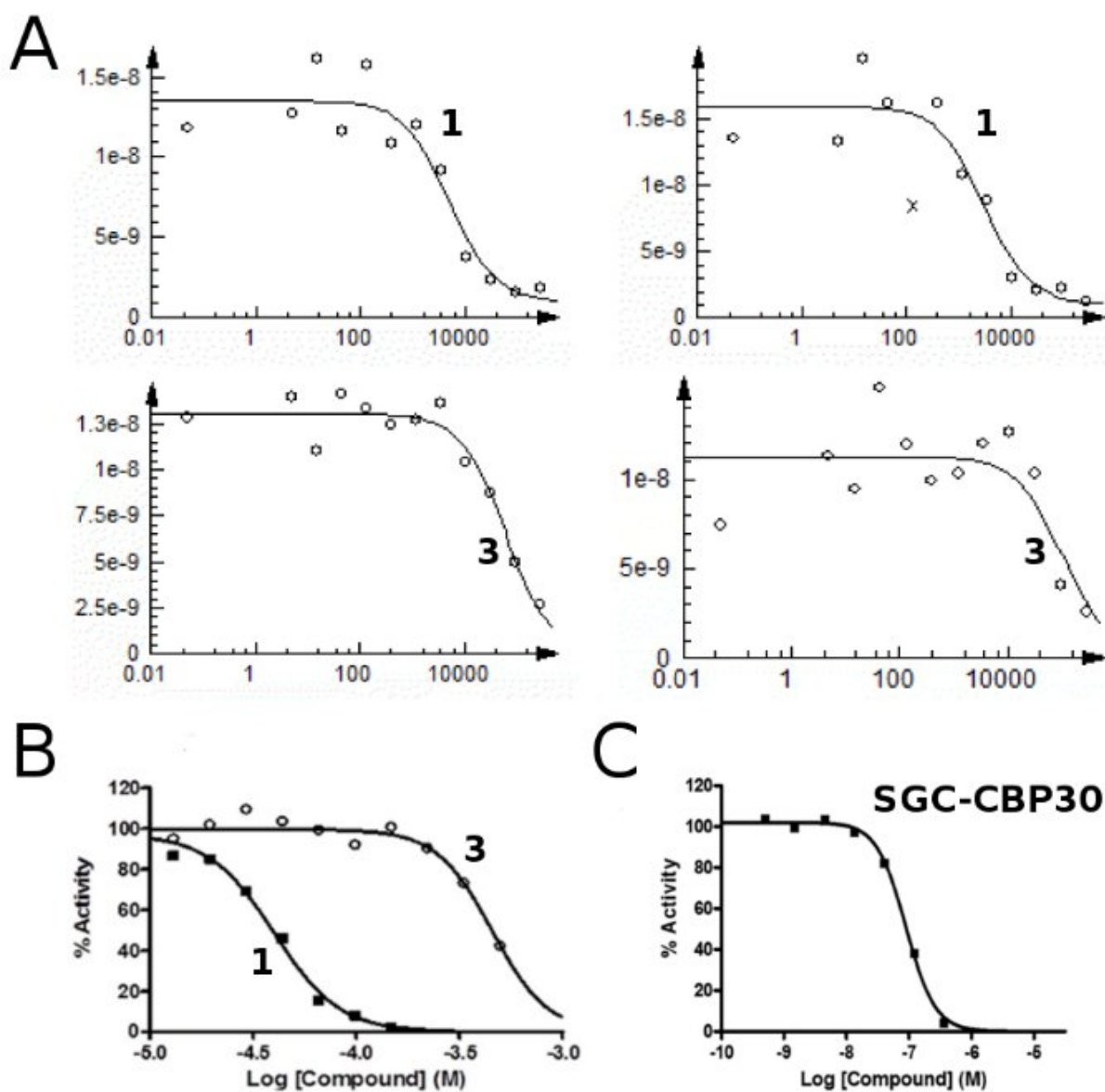
Figure S3. Competition binding assays for compounds **1** and **3**. (A) Dose-response curves in duplicates for the compounds **1** (top) and **3** (bottom) tested for binding to the CBP bromodomain in the BROMOscan competition binding assay. (B) Dose response curves for binding of hit fragments **1** (black squares) and **3** (open circles) to the CBP bromodomain as measured by the AlphaScreen assay using a biotinylated H4 peptide as competitive ligand at Reaction Biology Corporation. (C) SGC-CBP30 is used as a positive control; it consists of 36 heavy atoms and has a ligand efficiency of 0.31 kcal mol^{-1} HAC^{-1}.

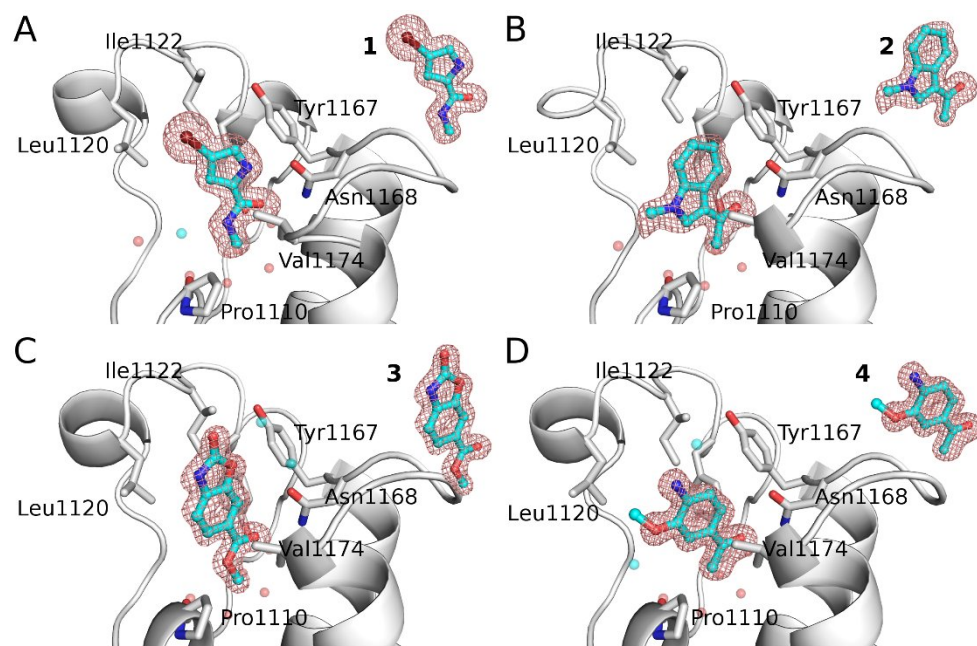**Figure S4**. Close view of binding mode of compounds **1**–**4**. Conserved water molecules in the binding pocket are shown as transparent spheres color coded as in Figure 3. Fo – Fc omit map for each compounds is shown in mesh contoured at 3 σ.
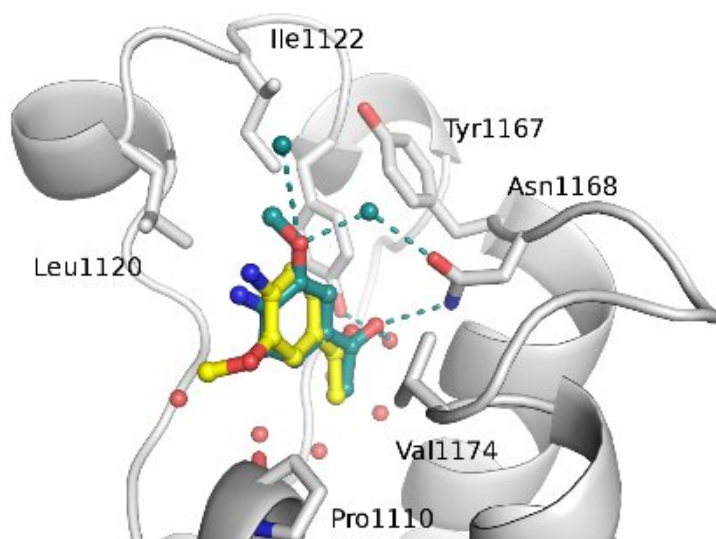
Figure S5. Comparison of the binding mode in the chain A of the crystal structure of the CBP/**4** complex (carbon atoms in deep teal, PDB code: 5MQG) with the binding pose predicted by docking with SEED [1,2] (carbon atoms in yellow). The oxygen atom of the methoxy group forms a water-mediated hydrogen bond with the side chain of the conserved asparagine. Conserved water molecules (red spheres) and crystallographic water molecules not used for docking (deep teal spheres) are shown. Hydrogen bonds are shown as deep teal dashed lines.
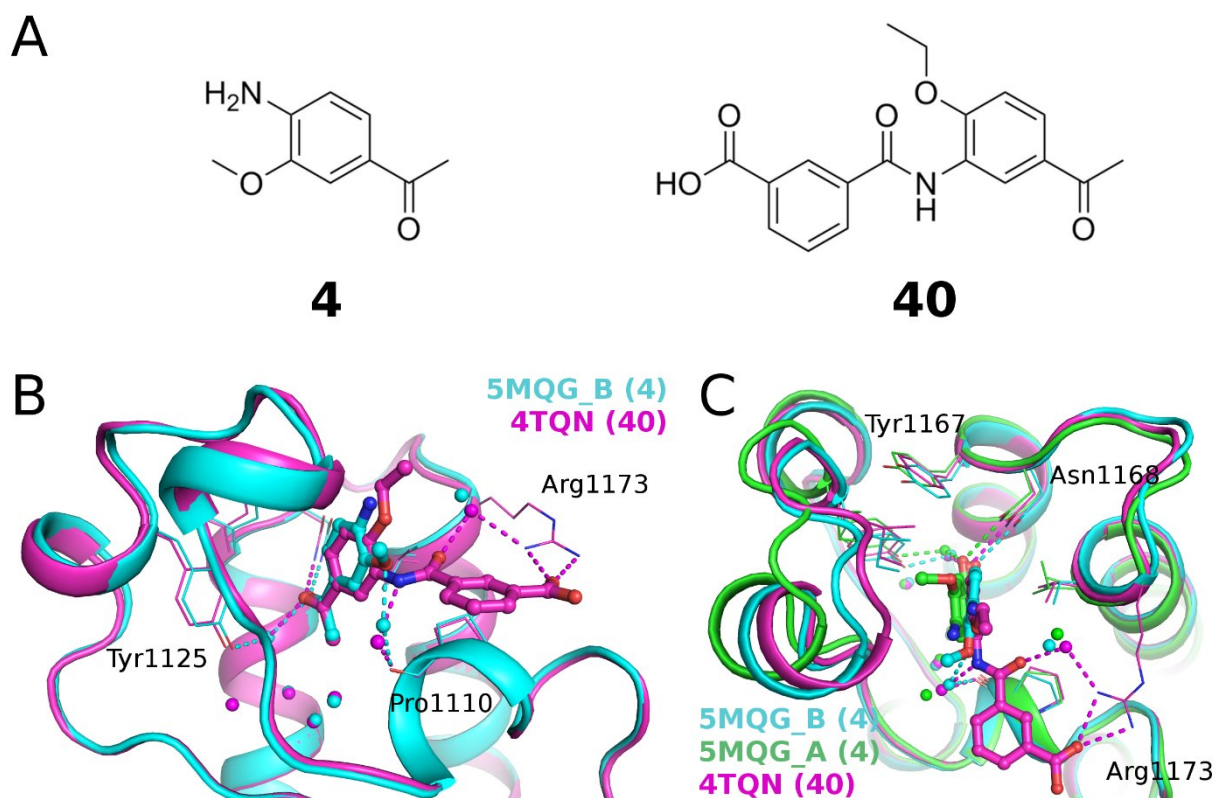
Figure S6. Comparison of the binding poses of compound **4** and the previously reported compound **40**. (A) Chemical structure of compounds **4** and **40**. (B) Comparison of the CBP/**4** complex (cyan) to the CBP/**40** complex (magenta, PDB code: 4TQN). Compound **4** recapitulates many polar interactions of the larger compound **40** (color-coded dashed lines), including a water-bridged hydrogen bond with the backbone oxygen of Pro1110. The smaller compound **4**, though, does not establishes any interaction with the side chain of Arg1173. (C) Comparison of the binding pose of compound **4** in the two chains of the asymmetric unit (cyan and green) and compound **40** (magenta).
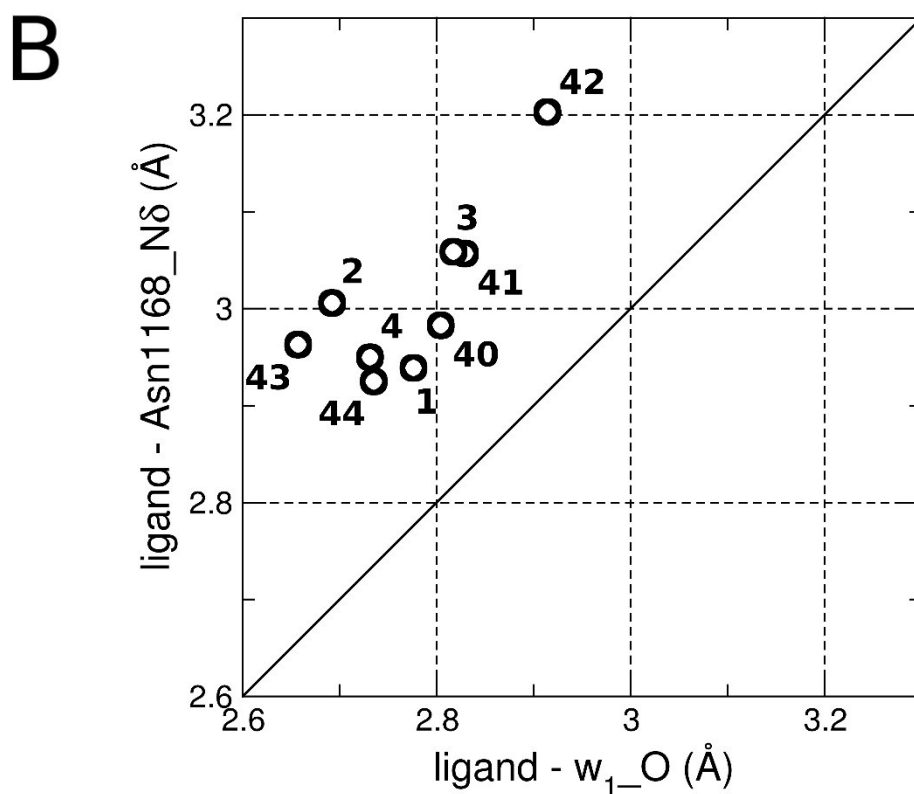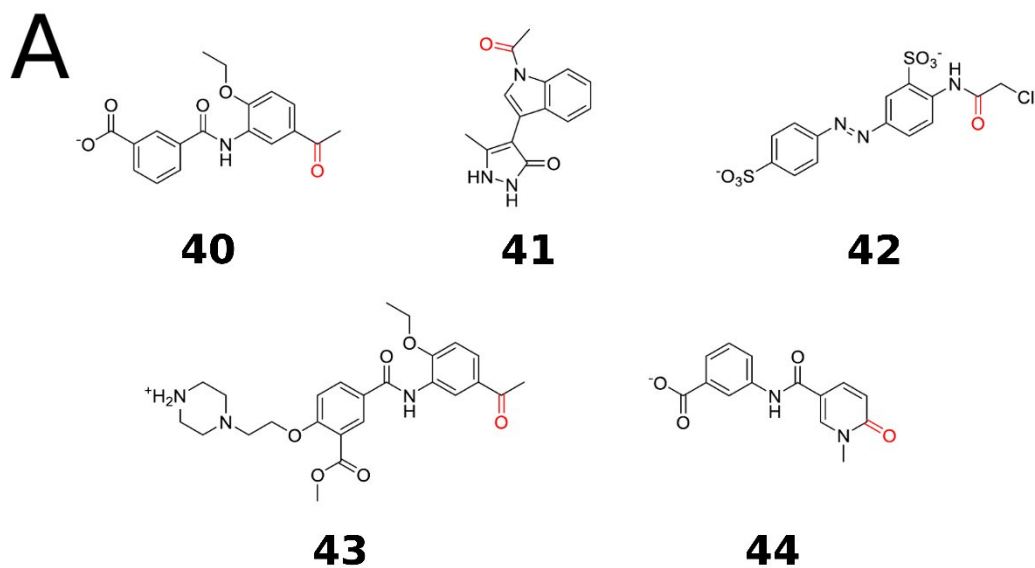
Figure S7. Analysis of the H-bonds of the group that mimics the acetyl of Kac. (A) 2D structures of compounds **40** (**10** in [26], PDB code: 4TQN), **41** (**1b** in [27], PDB code: 4TS8), **42** (PDB code: 5EIC), **43** (PDB code: 5ENG), and **44** (PDB code: 5EP7). The carbonyl group involved in the H-bond is in red. (B) Scatter plot of the distances involving the side chain nitrogen of Asn1168, with numbers indicating the ligand. In all cases the hydrogen bond distance with the conserved water w_1 that bridges to the Tyr1125 side chain is shorter than the distance to the Asn1168.

REFERENCES

1. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins*. 1999;37: 88-105.
2. Majeux N, Scarsi M, Caflisch A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins*. 2001;42(2): 256-268.
3. MacKerell AD, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*. 1998;102(18): 3586-3616.
4. MacKerell AD, Feig M, Brooks CL, 3rd. Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc*. 2004;126(3): 698–699.
5. Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010;31(4): 671-690.
6. Scarsi M, Apostolakis J, Caflisch A. Continuum electrostatic energies of macromolecules in aqueous solutions. *J Phys Chem A*. 1997;101(43): 8098-8106.
7. Zhao H, Huang D. Hydrogen bonding penalty upon ligand binding. *PLoS ONE*. 2011;6(6): e19923.
8. Zhao H, Gartenmann L, Dong J, Spiliotopoulos D, Caflisch A. Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg Med Chem Lett*. 2014;24(11): 2493-2496.
9. Huang D, Caflisch A. Library screening by fragment-based docking. *J Mol Recognit*. 2010;23(2): 183-193.
10. Zhao H, Caflisch A. Molecular dynamics in drug design. *Eur J Med Chem*. 2015;91: 4-14.
11. Filippakopoulos P, Picaud S, Mangos M, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*. 2012;149(1): 214-231.
12. Hwang T-L, Shaka AJ. Water suppression that works. Excitation sculpting using arbitrary waveforms and pulse field gradients. *J Magn Reson A*. 1995;112: 275-279.
13. Mayer M, Meyer B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem, Int Ed*. 1999;38: 1784-1788.
14. Carr HY, Purcell EM. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev*. 1954;94(3): 630-638.
15. Hay DA, Fedorov O, Martin S, et al. Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *J Am Chem Soc*. 2014;136(26): 9308-9319.
16. Philpott M, Yang J, Tumber T, et al. Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol bioSyst*. 2011;7(10): 2899-2908.
17. Quinn E, Wodicka L, Ciceri P, et al. BROMOscan - a high throughput, quantitative ligand binding platform identifies best-in-class bromodomain inhibitors from a screen of mature compounds targeting other protein classes. *Cancer Res*. 73. 2013:4238.
18. Kabsch W. Xds. *Acta Crystallogr D Biol Crystall*. 2010;66(Pt 2): 125-132.
19. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystall*. 2011;67(Pt 4): 235-242.
20. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystall*. 2013;69(Pt 7): 1204-1214.
21. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40(Pt 4): 658-674.
22. Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystall*. 2010;66(Pt 1): 22-25.
23. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystall*. 2010;66(Pt 4): 486-501.
24. Adams PD, Afonine PV, Bunkoczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystall*. 2010;66(Pt 2): 213-221.
25. Huang D, Rossini E, Steiner S, Caflisch A. Structured water molecules in the binding site of bromodomains can be displaced by cosolvent. *ChemMedChem*. 2014;9(3): 573-579.
26. Xu M, Unzue A, Dong J, Spiliotopoulos D, Nevado C, Caflisch A. Discovery of CREBBP bromodomain inhibitors by high-throughput docking and hit optimization guided by molecular dynamics. *J Med Chem*. 2016;59(4): 1340-1349.

27.   Unzue A, Zhao H, Lolli G, et al. The "gatekeeper" residue influences the mode of binding of acetyl indoles to bromodomains. *J Med Chem.* 2016;59(7): 3087-3097.