



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/110037/>

Version: Accepted Version

Article:

Sergio, Ruiz-Carmona, Schmidtke, Peter, Luque, F. Javier et al. (2017) Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nature Chemistry*. pp. 201-206. ISSN: 1755-4349

<https://doi.org/10.1038/nchem.2660>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

1 **Dynamic Undocking and the Quasi-Bound State as Tools for Drug Design**

2

3 Sergio Ruiz-Carmona,¹ Peter Schmidtke,² F. Javier Luque,¹ Lisa Baker,³ Natalia

4 Matassova,³ Ben Davis,³ Stephen Roughley,³ James Murray,³ Rod Hubbard,^{3,4} Xavier

5 Barril^{1,5,*}

6

7 ¹ Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Facultat de
8 Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.

9 ² Discngine, 33 rue du Fauburg Saint-Antoine, 75011 Paris, France

10 ³ Vernalis (R&D) Ltd, Granta Park, Cambridge, CB21 6GB, UK

11 ⁴ YSBL, University of York, Heslington, York, YO10 5DD, UK

12 ⁵ Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís
13 Companys 23, 08010 Barcelona, Spain.

14

15 * Send correspondence to: xbarril@ub.edu

16 There is a pressing need for new technologies that improve the efficacy and
17 efficiency of drug discovery. Structure-based methods have contributed towards
18 this goal but they focus on predicting the binding affinity of protein–ligand
19 complexes, which is notoriously difficult. We adopt an alternative approach that
20 evaluates structural, rather than thermodynamic, stability. Noting that bioactive
21 molecules present a static binding mode, we devised Dynamic Undocking (DUck), a
22 fast computational method to calculate the work necessary to reach a quasi-bound
23 state, where the ligand has just broken the most important native contact with the
24 receptor. This non-equilibrium property is surprisingly effective in virtual
25 screening because true ligands form more resilient interactions than decoys.
26 Notably, DUck is orthogonal to docking and other ‘thermodynamic’ methods. We
27 demonstrate the potential of the docking–undocking combination in a fragment
28 screening against the molecular chaperone and oncology target Hsp90, for which
29 we obtain novel chemotypes and a hit rate approaching 40%.

30
31
32
33
34
35
36
37
38
39
40

41 Structural stability is a fundamental property of protein–ligand complexes. Though
42 cases of dual binding modes have been reported,^{1,2} they are generally not dynamic,
43 or involve predominantly hydrophobic interactions,³ which lack directionality and
44 do not impose strict geometric constraints.⁴ By contrast, hydrogen bonds are ideal
45 to provide structural stability because they have sharp distance and angular
46 dependencies.⁴ Their contribution to the free energy of binding (ΔG_{bind}) is variable
47 but can be substantial.⁵ Importantly, they often act as anchoring points in protein–
48 ligand complexes, providing the minimal binding unit through one or a few
49 hydrogen bonds as demonstrated for fragment-sized ligands.^{6,7} We have
50 previously shown that certain hydrogen bonds present strong opposition to small
51 structural distortions and can act as kinetic traps because the local environment
52 hinders the transition from a direct hydrogen bond to a water-bridged
53 interaction.⁸ As an early unbinding event, rupture of the so-called water-shielded
54 hydrogen bonds can influence the whole dissociation process.^{8,9} Taken together,
55 these observations suggest that hydrogen bonds are the main determinants of
56 structural stability, and lead us to postulate that their resilience should provide
57 information about the binding potential of candidate ligands. Thus, we set out to
58 investigate whether the work required to disrupt intermolecular hydrogen bonds
59 can be used to predict ligand binding.

60

61 We will introduce DUCK, a simplified computational procedure to calculate the
62 work needed to break a key native contact, reaching a quasi-bound state (W_{QB}).
63 Then, we will show that active compounds are structurally stable and present
64 higher W_{QB} values than inactive ones. Finally, we demonstrate the use of this

65 property in virtual screening (VS) applications, showing that DUck complements
66 the thermodynamic perspective offered by existing methods.

67

68

69 **Results and Discussion**

70

71 Simplified simulation of the early dissociation stage

72 To assess the hypothesis, we have devised Dynamic Undocking (DUck)
73 simulations, where a key intermolecular hydrogen bond is pulled from an initial
74 distance of 2.5 Å (close contact) to 5.0 Å (broken contact). In order to focus on just
75 one specific hydrogen bond, we use model receptors comprising only the protein
76 residues that are within 6 Å of the given hydrogen bond (Figure 1A). The work
77 necessary to carry out the steering process is monitored, and we define the quasi-
78 bound (QB) state as the point along the simulation where the work profile
79 presents the highest value. W_{QB} is the work necessary to depart from the ideal
80 hydrogen bond configuration and reach the QB state (Figure 1B). Notably, this is a
81 non-equilibrium property, and there is no reason why it should correlate with any
82 measurement of binding affinity. What is more, as the unbound state is not
83 considered, W_{QB} cannot inform about the binding free energy. Instead, this
84 magnitude solely indicates if the interaction under investigation gives rise to a
85 (local) minimum in the free energy landscape and estimates the depth of said
86 minimum (Supplementary Figure 1).

87

88 Relationship between W_{QB} and binding affinity

89 As an initial proof of concept, we apply DUck to a set of 41 fragment-like ligands
90 (<300Da) of the cyclin dependent kinase 2 (CDK2) with known binding mode and
91 half maximal inhibitory concentration (IC_{50}) values. The hinge region of all kinases
92 is a hot spot for binding, where the protein backbone offers privileged hydrogen-
93 bonding opportunities.¹⁰ For CDK2, the central hydrogen-bond donor (NH of
94 Leu83) is the most conserved interaction site and was used to define the reaction
95 coordinate. W_{QB} presents only a weak correlation with binding affinity
96 (Supplementary Figure 2), but the distribution of W_{QB} values is clearly skewed
97 (Figure 2A and Supplementary Figure 3). Thus, 65% of weak binders ($IC_{50} > 1 \mu M$)
98 present W_{QB} values below 6 kcal/mol, while all strong binders ($IC_{50} < 1 \mu M$) pass
99 this threshold. Ligand 3FZ1,¹¹ is the clear exception as it presents an almost flat
100 dissociation profile ($W_{QB} = 0.12$ kcal/mol). This is explained by an unsuitably long
101 (3.4 Å) interaction with the hinge region, involving a methoxy group, which is a
102 poor hydrogen bond acceptor.⁴ Instead, this unusual ligand forms two charge-
103 reinforced hydrogen bonds with Lys33 and Asn132, from which it draws structural
104 stability (Supplementary Figure 4). This shows that some ligands can use
105 alternative or additional interaction points to attain structural stability, in which
106 case, DUck calculations (as currently implemented) may underestimate the cost of
107 breaking the native contacts.

108

109 To further examine the surprising relationship between binding affinity and W_{QB} ,
110 we use the bromodomain and extra-terminal (BET) BRD4-BD1 as additional test
111 system. The side-chain N of Asn140 is a well-known pharmacophoric point of this
112 epigenetic target,¹² and defines the key intermolecular hydrogen bond. Again, we
113 observe the same trend, i.e. higher W_{QB} for more potent ligands, but with a large

114 dispersion that blurs correlation (Supplementary Figures 3 and 5). Interestingly,
115 the lowest W_{QB} values (0, 1.1 and 1.7 kcal/mol) correspond to three kinase
116 inhibitors with off-target activity for the BRD4-BD1.¹³ Thus, achieving potency in
117 the absence of a robust anchoring interaction is possible, but rare, which suggests
118 that it is an ineffective strategy.

119

120 DUck is very effective in virtual screening

121 We then assess whether the approach can be used in virtual ligand screening by
122 testing the ability of DUck to distinguish true CDK2 ligands from a set of carefully
123 selected decoys¹⁴ for which we had generated binding modes by docking. The
124 distribution of W_{QB} is strikingly different from the active set, with 61% of
125 molecules presenting values below 2 kcal/mol and 49% below 1 kcal/mol (Figure
126 2A). This indicates that, in spite of forming the key hydrogen bond, this interaction
127 is labile for most of the docking decoys, which would translate to an unstable
128 binding mode. We therefore propose that W_{QB} can distinguish true ligands from
129 inactive molecules, as shown in the receiver operating characteristics (ROC) curves
130 (Figure 2B). To demonstrate the wider applicability of the method, we conducted
131 similar experiments with the adenosine A2A receptor (AA2R) and Trypsin, as
132 representatives of G protein-coupled receptors (GPCR) and serine proteases,
133 respectively (Figure 2B). Together with kinases (such as CDK2) these protein
134 families include a large part of the current and investigational drug targets.¹⁵ The
135 key hydrogen bonds tracked by the DUck simulations involve the side-chain
136 carbonyl of Asn253, in the case of AA2R, and the carboxylic acid of Asp189, for
137 Trypsin. As shown in Figure 2B, the results for these systems are even better than
138 for CDK2, demonstrating that DUck is surprisingly effective in virtual screening.

139 Importantly, the performance improves consistently as sampling increases, but
140 good enrichments can be obtained with as little as 2 DUck runs per ligand
141 (Supplementary Figure 6).

142

143 DUck is orthogonal to existing methods

144 These results position DUck as a new method for virtual screening. But, as it aims
145 to predict a property that is fundamentally different from thermodynamic stability,
146 we investigate its complementarity with molecular docking, a method with a long
147 and successful history of application in virtual screening.^{16,17} Using the rDock
148 software,¹⁸ we find that docking scores have no correlation with W_{QB} , and good
149 docking scorers are nearly as likely to present a low resistance to dissociation as
150 the rest of the decoys (Figures 2C, 2D and Supplementary Figure 7). As such,
151 molecular docking and dynamic undocking can be considered orthogonal (i.e.
152 perfectly complementary) and the intersection between both techniques defines a
153 region highly enriched in true ligands. We have also performed extensive
154 calculations with other virtual screening tools (Glide docking, MMPBSA and
155 MMGBSA re-scoring). The results, summarised in Supplementary Figures 8 and 9,
156 confirm that DUck is complementary to all of them. In fact, as we obtain low W_{QB}
157 values for many decoys with good scores by all other methods, DUck post-filtering
158 delivers several fold improvement even when applied to a consensus list by two
159 independent ‘thermodynamic’ approaches (Figures 2E, 2F and Supplementary
160 Figure 10). These results support the idea that structural stability of the binding
161 mode, just like good chemical complementarity, is a necessary – but not sufficient –
162 condition for binding. By imposing both conditions simultaneously, we can
163 multiply the effectiveness of structure-based virtual screening. At the same time,

164 using W_{QB} as a post-docking filter means that only the best-scoring subset of the
165 virtual chemical collection needs to be reassessed by DUck simulations, thus
166 improving computational efficiency.

167

168 Fragment discovery with in tandem docking-undocking calculations.

169 To demonstrate the power of the docking-undocking combination, we have
170 applied the method prospectively for the identification of small molecules that
171 bind the molecular chaperone, Heat Shock Protein 90KDa (Hsp90). This oncology
172 target has been a test-bed and paradigm in fragment and structure-based drug
173 design.¹⁹ With hundreds of Hsp90-ligand complexes deposited in the Protein Data
174 Bank (PDB), discovery of novel chemotypes is very challenging. We focused on
175 fragment-like molecules, as this may be the most efficient way to discover new
176 leads and to generate scaffold-hopping ideas.^{20,21} A collection of 280000 fragment-
177 sized molecules was docked to the ATP binding site of Hsp90. A diverse set of 139
178 molecules from the best 450 (top 0.16%) was then selected and each one was
179 subjected to 100 DUck runs to obtain fully converged W_{QB} values (note that fewer
180 DUck runs would have given similar results (Supplementary Figure 11)). The
181 distribution of W_{QB} values (Figure 3A, Supplementary Figure 12) shows that even
182 at the upper limit of the docking score distribution a large proportion of putative
183 ligands present low resistance to dissociation, with 32%, 50% and 80% presenting
184 W_{QB} below 3, 4 and 6 kcal/mol, respectively. We purchased all the molecules from
185 the high stability set ($W_{QB} > 6$ kcal/mol) that were available (n=21). They were
186 tested using three different ligand-observed Nuclear Magnetic Resonance (NMR)
187 experiments, in the absence or presence of a known competitor to confirm that
188 fragment hits bind at the target site.¹⁹ Eight out of the 21 molecules (38%) were

189 confirmed as true hits (Table 1). Crucially, for the same system and screening
190 method, the hit rate obtained with a general fragment screening library is 4.4%.²²
191 Therefore, the DUck-based virtual screening increases the efficiency by nearly an
192 order of magnitude. This is similar to optimal virtual fragment screening results
193 reported for other systems.²³ In order to better assess the contribution of DUck to
194 the success rate, we also purchased and collected data for 15 molecules from the
195 medium stability set (W_{QB} between 3 and 6 kcal/mol) and 11 from the low stability
196 set ($W_{QB} < 3$ kcal/mol). Only one molecule from these sets was a hit and,
197 importantly, its W_{QB} value is very close to the upper threshold (5.6 kcal/mol). This
198 confirms that DUck false negatives (i.e. active molecules with low W_{QB}) are rare, an
199 ideal property for a screening method. Hit rates for the three categories are
200 summarized in Figure 3B.

201

202 To assess the value of the hits as starting points, we have compared their chemical
203 structures to existing Hsp90 ligands, finding low similarity in all cases (Table 1).
204 Binding mode determination and analysis of the main interactions that define the
205 chemical scaffold offers a more precise assessment of their novelty. Crystal
206 structures for 3 of the fragment hits were determined by X-ray crystallography
207 (Figure 4 and Supplementary Figure 13). This confirmed that the docking pose
208 used as starting position for the DUck experiments was correct, particularly
209 regarding the key interaction that was being monitored (side-chain of Asp93).
210 Compound **1** is the most potent fragment hit (dissociation constant $K_D=77\mu\text{M}$) and
211 has a ligand efficiency (LE) of 0.33 kcal/mol per non-hydrogen atom, similar to
212 other Hsp90 fragment hits that have been evolved into very efficient lead
213 compounds.²⁴ Many 2-aminopyrimidines have been described as Hsp90 ligands,¹⁹

214 confirming the potential of the fragment hit, but the relative lack of novelty would
215 advise against using this fragment as starting point at this stage. Compound **2** is
216 less potent ($K_D=320\mu\text{M}$) but equally efficient ($LE=0.32$) by virtue of having fewer
217 atoms. In this case, the key interaction with Asp93 is mediated by an aminothiazole
218 moiety, which is unprecedented and would constitute a good starting point to
219 develop new chemical entities. Compound **3** ($K_D=700\mu\text{M}$; $LE=0.25$) belongs to the
220 well-known family of resorcinol inhibitors, which includes the clinical candidate
221 NVP-AUY922,¹⁹ but provides an interesting example of scaffold hopping, where the
222 oxime acts a bioisosteric replacement of the five-membered rings included as core
223 scaffold in the patents. Compounds **4**, **5** and **6** also represent completely novel
224 starting points, as their scaffold is unique amongst Hsp90 inhibitors. The binding
225 mode could not be confirmed experimentally, but is likely correct because two
226 independent methods deemed the molecules active based on the predicted
227 geometry (Their predicted binding modes can be found in Supplementary Figure
228 14).

229

230 **Conclusions**

231

232 In summary, we have demonstrated that the concept of structural stability can be
233 used very effectively in structure-based drug design, complementing the standard
234 focus on binding free energy. Hydrogen-bonding groups in the active site are
235 privileged structures to fix the ligand in place, particularly when they act as
236 binding hot spots and can form water-shielded hydrogen bonds.⁸ The work needed
237 to break such interactions (W_{QB}) is very useful to detect true ligands even though it
238 is a non-equilibrium property that is not expected to correlate with ΔG_{bind} . This

239 intriguing fact may reflect the nature of proteins, which have been designed to
240 bind their natural ligands not only with high affinity and selectivity, but also
241 forming structurally stable complexes. Thus, it will be important to test the
242 approach on other types of supramolecular assemblies. Dynamic Undocking
243 (DUck), a particular implementation of steered molecular dynamics, allows us to
244 calculate W_{QB} in a very efficient manner. DUck can be used in combination with
245 existing ‘thermodynamic’ approaches to multiply their effectiveness. The docking-
246 undocking combination has proven particularly useful for virtual fragment
247 screening, delivering novel, diverse and suitable starting points with a hit rate of
248 38%. At present, we focus on a single key hydrogen bond to estimate W_{QB} , which
249 requires previous knowledge and has a critical impact on the outcome. Future
250 investigations should address the extension of the method to multiple sites and
251 other interaction types to improve performance and avoid reliance on extrinsic
252 decisions. DUck inherits the intrinsic limitations of structure-based methods (e.g.
253 protein flexibility, quality of the force-field) and may have some of its own (e.g.
254 long range effects, steering conditions). Further tests will reveal its true potential,
255 but considering that it is orthogonal to existing methods and computationally very
256 efficient, we expect that it will be rapidly adopted by the structure-based drug
257 design community and adapted to other biotechnological applications involving
258 non-covalent complexes.

259

260

261 **METHODS**

262

263 **Dynamic Undocking**

264 Dynamic Undocking (DUck) is a particular type of Steered Molecular Dynamics
265 (SMD),²⁵ where we force the rupture of an intermolecular hydrogen bond formed
266 between a pre-defined interaction point in the receptor and a complementary
267 atom in the ligand. Additionally, we use a model receptor that includes only the
268 minimal subset of the protein necessary to preserve the local environment around
269 the hydrogen bond that is being monitored. This transformation minimizes the
270 influence of peripheral interactions, thus simplifying the dissociation pathway and
271 facilitating convergence (Supplementary Figure 15). As an added bonus, it speeds
272 up the calculations by a factor of 5 (Supplementary Table 2). The first and essential
273 step is to identify an atom of reference in the protein, which must form a hydrogen
274 bond with all (or most) known ligands. For well-known systems, like the ones used
275 here, it can be identified from a structural superimposition of all the available
276 protein-ligand complexes. On novel binding sites, it may be identified with a
277 quantitative hot spot identification method.²⁶ Then, the model receptor is
278 generated from a representative 3D structure of the protein by selecting all
279 residues with at least one atom within 6 Å of the atom of reference. The selection is
280 visually inspected and, if needed, additional residues that are deemed necessary to
281 preserve the local environment are included in the selection. Unselected residues
282 are eliminated and truncated side chains are acetylated or N-methylated, as
283 needed. Interstitial water molecules, if present, are preserved. The PDB codes,
284 reference interaction points and the list of protein residues and water molecules
285 for each system are listed in Supplementary Table 3. Given the model receptor

286 (protein chunk) and a set of ligands properly oriented (docking poses or
287 superimposed X-ray geometries), a MOE²⁷ SVL script developed in house
288 automatically performs the following steps: 1) Calculates AM1-BCC charges for the
289 ligand.²⁸ 2) Assigns parm@Frosst²⁹ atom types and non-bonded parameters to the
290 ligand. 3) Identifies the ligand atom that is hydrogen-bonded to the protein's
291 reference atom (based on distance and type). 4) Writes input and execution files to
292 carry out the MD simulations with AMBER³⁰. 5) Calls AMBER's tLeap to generate
293 valid topology and coordinate files for each individual receptor-ligand complex.
294 For the protein, the AMBER force field 99SB is used. Each system is placed in a
295 cuboid box spanning at least 12 Å more than the furthest atom in each direction.
296 The box is then filled with TIP3 water molecules to create periodic boundary
297 conditions. When needed, Na⁺ or Cl⁻ ions are added to force the neutrality of the
298 whole system. MD simulation conditions (where non-default) are as follows: 1) At
299 all stages, harmonic restraints with a force constant of 1 kcal/mol·Å² are placed on
300 all non-hydrogen atoms of the receptor to prevent structural changes. 2)
301 Spontaneous rupture of the key hydrogen bond during non-steered simulations is
302 prevented with a gradual restraint for distances beyond 3 Å (parabolic with k=1
303 kcal/mol·Å² between 3Å and 4Å and linear with k=10 kcal/mol·Å beyond 4 Å). 3)
304 All equilibration and simulation steps were run using a Langevin thermostat with a
305 collision frequency of 4 ps⁻¹ and the cutoff for non-bonded interactions was set to
306 9Å. 4) Bonds involving hydrogen are constrained using SHAKE.³¹ In order to
307 equilibrate the system the following steps are executed: 1) Energy minimization
308 for 1000 cycles. 2) Assignment of random velocities at 100K and gradual warming
309 to 300K for 400 ps in the NVT ensemble. 3) Equilibration of the system for 1 ns in
310 the NPT ensemble (1 atm, 300K). At this stage, the first SMD simulations can be

311 executed. We run two SMDs from the same restart file, but at different
312 temperatures (300K and 325K) to ensure that the trajectories proceed differently.
313 The SMD lasts 500 ps, during which time the distance between the key hydrogen
314 bonds is steered from 2.5 Å to 5.0 Å (constant velocity of 5 Å/ns) with a spring
315 constant of 50 kcal/mol·Å². We have tested slower velocities and the results are
316 essentially unchanged (Supplementary Figure 16). The spring constant had little
317 influence and on a limited test set we obtained essentially identical results in the
318 range k=10 kcal/mol·Å² to k=1000 kcal/mol·Å². We have also investigated the
319 importance of the specific reaction coordinate by using the closest contact
320 between CDK2 Leu83:O and the ligand (instead of Leu83:N). The W_{QB} values
321 obtained with these different atoms of reference (located only 3 Å apart) present a
322 high correlation ($r^2=0.75$; Supplementary Figure 17). By contrast, when the atoms
323 of reference involve completely different part of the ligand, the results are
324 uncorrelated (Supplementary Figure 18). To generate diverse starting points for
325 SMD trajectories, we perform 1ns unbiased MD simulation and repeat the process
326 as many times as desired (e.g. 50ns unbiased MD simulations are needed to
327 execute 100 SMD trajectories). All simulations were performed with Amber 12
328 adapted for running in GPUs and executed either in-house with NVIDIA GeForce
329 TITAN X GPUs or at the Barcelona Supercomputing Center using NVIDIA Tesla
330 M2090 GPUs. The simulations took 24 minutes (unbiased MD) or 30 minutes
331 (SMD) of wallclock time per nanosecond (average values for the systems tested on
332 the TITAN GPUs). Work profiles outputted by the SMD simulations are processed
333 as explained in the main text to obtain W_{QB} values. Various methods could be used
334 to obtain free energies from the SMD work, but they have strict convergence
335 requirements, are computationally much more expensive and the results are only

336 valid if the reaction coordinate is mechanistically correct.²⁵ Instead, we simply
337 assume that W_{QB} is an upper limit to the equivalent magnitude in free energy
338 (ΔG_{QB}). In order to get as close as possible to ΔG_{QB} , we run multiple SMD replicas
339 and take the overall lower W_{QB} as the representative value. Note that we have used
340 very conservative settings, favouring sampling over computational efficiency.
341 Based on convergence analysis (Supplementary Figures 6 and 11) and other tests,
342 we propose the protocol shown in Supplementary Figure 19 for virtual screening.
343 Less than one GPU hour per ligand would be necessary to discard approximately
344 80% of candidate ligands and produce a reasonable estimate of W_{QB} for the
345 remaining ones. By comparison, a high-throughput implementation of MM-PBSA (1
346 ns of sampling) would require at least 3 GPU hours plus 20 CPU minutes per
347 ligand.

348

349 **Hsp90 virtual screening**

350 A collection of 280000 purchasable fragment-sized molecules (<250 Da), were
351 docked to the ATP binding site of Hsp90 with an optimized protocol, where the key
352 hydrogen bond with Asp93 is enforced.¹⁸ Next, we grouped the 1000 top scoring
353 molecules into 400 clusters based on chemical similarity and visually inspected the
354 top-scoring molecule within each cluster to select 139 molecules that were
355 subjected to DUCK simulations. Docking score was the main selection criterion,
356 with 90 molecules originating from the top 200 and all of them within the top 450.
357 Additional criteria included high predicted aqueous solubility and chemical
358 diversity. The selected molecules were subjected to 100 DUCK calculations. We
359 divided the molecules in three categories according to their resistance to
360 dissociation: weak ($W_{QB} < 3$; N=44; 32%), medium ($3 < W_{QB} < 6$; N=67; 48%) and

361 strong ($W_{QB} > 6$; $N=28$; 20%). We tested all the molecules that we could buy from
362 the strong set. For comparison, we also purchased and tested 15 molecules of
363 medium and 11 from the low stability sets. The chemical structures of the 47
364 compounds are shown in Supplementary Figure 12.

365

366 **Screening by NMR**

367 Identification of compounds which bind to the ATP site of Hsp90 α was performed
368 as described previously.^{32,33} Briefly, a number of 1D ^1H NMR experiments (STD,
369 water-LOGSY, relaxation filtered) were used to identify interactions between
370 compounds and the protein; a potent competitor (PU3) was then added in order to
371 block the ATP binding site. Compounds which bound and were then displaced
372 were identified as interacting specifically with the protein.³⁴ Molecules active in all
373 experiments were considered *bona fide* hits, while those giving a positive response
374 in one or two experiments were considered unconfirmed hits because changes in
375 NMR signal are not necessarily related to binding. All NMR experiments were
376 performed on a BrukerAvIII HD 600 MHz NMR spectrometer at 298K; pulse
377 sequences included an excitation sculpting module in order to suppress bulk
378 water. Samples contained 500 μM ligand and 10 μM Hsp90 α in 20mM tris pH 7.5,
379 50mM NaCl 1mM freshly prepared DTT and contained 10% D_2O .

380

381 **X-Ray crystallographic studies**

382 Protein was produced and crystallized as previously described.³⁵ For the
383 successful crystals, data were collected at 100K on an in-house Bruker D8 Venture
384 TXS Generator with a Bruker Photo 100 detector and were subsequently
385 processed using SAINT & SADABS. The crystals belong to the space groups I222.

386 The structures were solved by molecular replacement using a previously solved
387 Hsp90 α protein model (PDB code: 1UY6; PU3 ligand and solvent removed) and the
388 program AMoRe.³⁶ Twenty cycles of rigid-body then restrained refinement were
389 carried out using the refinement program REFMAC5³⁷ followed by model building
390 and solvent addition using the molecular graphics program COOT.³⁸ The progress
391 of the refinement was assessed using R_{free} and the conventional R factor. Once
392 refinement was completed the structures were validated using various programs
393 from the CCP4i package.³⁹ Full data collection and refinement statistics are
394 presented in Supplementary Table 4.

395

396 Methodological details concerning the creation of the datasets, molecular docking,
397 MMPBSA and MMGBSA calculations, and surface plasmon resonance experiments
398 are provided as Supplementary Information.

399

400

401

402

403 **ACKNOWLEDGEMENTS**

404 We thank Carles Galdeano for helpful discussions and manuscript revision. We
405 thank the Barcelona Supercomputing Center for access to computational
406 resources. This work was financed by the Spanish Ministerio de Economía
407 (SAF2012-33481, SAF2015-68749-R), the Catalan government (2014 SGR 1189)
408 and the ICREA Academia program (FJL).

409 Correspondence and requests for materials should be addressed to X. B.

410

411 **Author contributions**

412

413 F.J.L, R.H. and X.B. conceived the overall strategy of the study. P.S. and X.B.
414 conceived and implemented the DUck approach. S.R.C. and X.B. carried out and
415 analysed all computational work. B.D. performed NMR experiments. L.B.
416 performed X-ray crystallography experiments. N.M. performed SPR experiments.
417 All authors discussed the results, designed experiments and wrote the paper.

418

419

420 **Code Availability statement**

421

422 A step-by-step guide on how to prepare, execute and analyse DUck simulations,
423 along with scripts used to automate the process can be found on the following
424 website: <http://www.ub.edu/bl/undocking>

425

426

427

428 **REFERENCES**

429

- 430 1. Kuhnert, M. *et al.* Tracing Binding Modes in Hit-to-Lead Optimization:
431 Chameleon-Like Poses of Aspartic Protease Inhibitors. *Angew. Chem. Int. Ed.*
432 *Engl.* **54**, 2849–2853 (2015).
- 433 2. Krohn, A., Redshaw, S., Ritchie, J. C., Graves, B. J. & Hatada, M. H. Novel
434 binding mode of highly potent HIV-proteinase inhibitors incorporating the
435 (R)-hydroxyethylamine isostere. *J. Med. Chem.* **34**, 3340–3342 (1991).
- 436 3. Smith, L. J., Gunsteren, W. F. Van & Allison, J. R. Multiple binding modes for
437 palmitate to barley lipid transfer protein facilitated by the presence of
438 proline 12. *Protein Sci.* **22**, 56–64 (2013).
- 439 4. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist's Guide to Molecular
440 Interactions. *J. Med. Chem.* 5061–5084 (2010). doi:10.1021/jm100112j
- 441 5. Klebe, G. Applying thermodynamic profiling in lead finding and optimization.
442 *Nat. Rev. Drug Discov.* **14**, 95–110 (2015).

- 443 6. Ferenczy, G. G. & Keserú, G. M. Thermodynamics of fragment binding. *J.*
444 *Chem. Inf. Model.* **52**, 1039–45 (2012).
- 445 7. Kozakov, D. *et al.* Ligand deconstruction: Why some fragment binding
446 positions are conserved and others are not. *Proc. Natl. Acad. Sci. U. S. A.* **112**,
447 E2585-2594 (2015).
- 448 8. Schmidtke, P., Luque, F. J., Murray, J. B. & Barril, X. Shielded hydrogen bonds
449 as structural determinants of binding kinetics: application in drug design. *J.*
450 *Am. Chem. Soc.* **133**, 18903–10 (2011).
- 451 9. Colizzi, F., Perozzo, R., Scapozza, L., Recanatini, M. & Cavalli, A. Single-
452 molecule pulling simulations can discern active from inactive enzyme
453 inhibitors. *J. Am. Chem. Soc.* **132**, 7361–71 (2010).
- 454 10. Noble, M. E. M., Endicott, J. a & Johnson, L. N. Protein kinase inhibitors:
455 insights into drug design from structure. *Science* **303**, 1800–5 (2004).
- 456 11. Anderson, D. R. *et al.* Benzothiophene inhibitors of MK2. Part 1: structure-
457 activity relationships, assessments of selectivity and cellular potency. *Bioorg.*
458 *Med. Chem. Lett.* **19**, 4878–81 (2009).
- 459 12. Filippakopoulos, P. & Knapp, S. Targeting bromodomains: epigenetic readers
460 of lysine acetylation. *Nat. Rev. Drug Discov.* **13**, 337–356 (2014).
- 461 13. Ember, S. W. J. *et al.* Acetyl-lysine binding site of bromodomain-containing
462 protein 4 (BRD4) interacts with diverse kinase inhibitors. *ACS Chem. Biol.* **9**,
463 1160–1171 (2014).
- 464 14. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful
465 decoys, enhanced (DUD-E): Better ligands and decoys for better
466 benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
- 467 15. Rask-Andersen, M., Almén, M. S. & Schiöth, H. B. Trends in the exploitation of
468 novel drug targets. *Nat. Rev. Drug Discov.* **10**, 579–90 (2011).
- 469 16. Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432**, 862–5
470 (2004).
- 471 17. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms.
472 *Annu. Rev. Biophys. Biomol. Struct.* **32**, 335–73 (2003).
- 473 18. Ruiz-Carmona, S. *et al.* rDock: A Fast, Versatile and Open Source Program for
474 Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **10**,
475 e1003571 (2014).
- 476 19. Roughley, S., Wright, L., Brough, P., Massey, A. & Hubbard, R. E. Hsp90
477 inhibitors and drugs from fragment and virtual screening. *Top. Curr. Chem.*
478 **317**, 61–82 (2012).
- 479 20. Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic
480 advances and lessons learned. *Nat. Rev. Drug Discov.* **6**, 211–9 (2007).
- 481 21. Joseph-McCarthy, D., Campbell, A. J., Kern, G. & Moustakas, D. Fragment-
482 based lead discovery and design. *J. Chem. Inf. Model.* **54**, 693–704 (2014).
- 483 22. Chen, I.-J. & Hubbard, R. E. Lessons for fragment library design: analysis of

- 484 output from multiple screening campaigns. *J. Comput. Aided. Mol. Des.* 603–
485 620 (2009). doi:10.1007/s10822-009-9280-5
- 486 23. Teotico, D. G. *et al.* Docking for fragment inhibitors of AmpC beta-lactamase.
487 *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7455–7460 (2009).
- 488 24. Murray, C. W. *et al.* Fragment-based drug discovery applied to Hsp90.
489 Discovery of two lead series with high ligand efficiency. *J. Med. Chem.* **53**,
490 5942–55 (2010).
- 491 25. Chipot, C. & Pohorille, A. *Free Energy Calculations Theory and Applications in*
492 *Chemistry and Biology. Free Energy Calculations Theory and Applications in*
493 *Chemistry and Biology* (Springer Berlin Heidelberg, 2007).
- 494 26. Alvarez-Garcia, D. & Barril, X. Molecular simulations with solvent
495 competition quantify water displaceability and provide accurate interaction
496 maps of protein binding sites. *J. Med. Chem.* **57**, 8530–9 (2014).
- 497 27. Chemical Computing Group Inc. Molecular Operating Environment (MOE),
498 2014.09. (2015).
- 499 28. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality
500 atomic charges. AM1-BCC model: II. Parameterization and validation. *J.*
501 *Comput. Chem.* **23**, 1623–41 (2002).
- 502 29. Bayly, C. I., McKay, D. & Truchon, J. F. An Informal AMBER Small Molecule
503 Force Field: parm@Frosst. (2011).
- 504 30. Case, D. A. *et al.* AMBER 12. (2012).
- 505 31. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the
506 cartesian equations of motion of a system with constraints: molecular
507 dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
- 508 32. Brough, P. a *et al.* Combining hit identification strategies: fragment-based
509 and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine
510 inhibitors of the Hsp90 molecular chaperone. *J. Med. Chem.* **52**, 4794–809
511 (2009).
- 512 33. Baurin, N. *et al.* Design and characterization of libraries of molecular
513 fragments for use in NMR screening against protein targets. *J. Chem. Inf.*
514 *Comput. Sci.* **44**, 2157–2166 (2004).
- 515 34. Davis, B. in *Protein-Ligand Interactions: Methods and Applications* (eds.
516 Williams, M. A. & Daviter, T.) **1008**, 389–413 (Springer Science+Business
517 Media, 2013).
- 518 35. Wright, L. *et al.* Structure-activity relationships in purine-based inhibitor
519 binding to HSP90 isoforms. *Chem. Biol.* **11**, 775–85 (2004).
- 520 36. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta*
521 *Crystallogr. Sect. A Found. Crystallogr.* **50**, 157–163 (1994).
- 522 37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular
523 structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol.*
524 *Crystallogr.* **53**, 240–255 (1997).

- 525 38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics.
526 *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- 527 39. Collaborative, C. P. & others. The CCP4 suite: programs for protein
528 crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760 (1994).
529

530 **TABLES**

531

532 **Table 1.** Summary of results for the 9 Hsp90 NMR Class 1 hits. Chemical structures
533 of all compounds are shown in Supplementary Table 1.

534

ID	MW	Docking	DUck	SPR Kd (mM)	PDB Sim ^b	ChEMBL Sim. ^b
		Score (Rank ^a)	Score (Rank ^a)			
1*	248.7	-25.0 (79)	9.1 (10)	77	2XDX (0.37)	CHEMBL 1340447 (0.44)
2*	221.3	-25.0 (73)	8.2 (11)	320	2WI6 (0.29)	CHEMBL 1536318 (0.54)
3*	230.2	-26.7 (19)	11.3 (1)	700	4EFU (0.32)	CHEMBL 1458840 (0.51)
4	240.3	-26.4 (22)	7.4 (16)	730	3WHA (0.29)	CHEMBL 1542436 (0.37)
5	165.2	-23.8 (128)	8.1 (12)	-	4EFT (0.27)	CHEMBL 1313412 (0.28)
6	206.3	-23.3 (138)	9.5 (5)	-	3HHU (0.42)	CHEMBL 2103879 (0.42)
7	236.3	-25.4 (51)	7.8 (15)	-	3B24 (0.31)	CHEMBL 1375884 (0.36)
8	224.7	-25.3 (58)	7.0 (22)	-	2XDX (0.35)	CHEMBL 1383799 (0.37)
22	237.3	-28.3 (2)	5.6 (33)	-	300I (0.27)	CHEMBL 1834092 (0.33)

535

536 *Xray structure solved ^aPosition within the list

537 of 149 molecules that were evaluated with

538 DUck. ^bHsp90 structure in the PDB or

539 compound with Hsp90 activity in ChEMBL (as

540 of 23/03/2016) with the closest similarity to

541 the fragment hit. Similarity (values in

542 parentheses) was calculated with Open Babel

543 using the FP2 fingerprint.

544

545

546 **FIGURE CAPTIONS**

547

548

549 **Figure 1.** Calculation of W_{QB} . **a.** The receptor is idealized as a model system
550 containing only the local environment around a key intermolecular hydrogen
551 bond. **b.** Representative work profiles obtained from dynamic undocking
552 simulations for a strong (black) and a weak (grey) ligand. The quasi-bound state is
553 defined as the point with the highest energy relative to the ideal hydrogen bond
554 geometry.

555

556

557 **Figure 2.** Application of the quasi-bound approximation to ligand ranking. **a.**
558 Distribution of W_{QB} values of potent CDK2 ligands ($IC_{50} < 1\mu M$; dark grey), weak
559 CDK2 ligands ($IC_{50} > 1\mu M$; light grey) and non-binding decoys (black). Points
560 indicate population values, from which the smooth lines are extrapolated. **b.** ROC
561 curves for the CDK2 (black), A2AR (red) and Trypsin (green) DUD sets. Plotted
562 results correspond to 2 DUck runs per ligand. AUC values are shown in
563 Supplementary Figure 6. **c.** Docking score vs. W_{QB} values for active (red) and
564 inactive (black or gray) compounds in the CDK2 retrospective virtual screening
565 dataset. The quadrant in orange highlights the area corresponding to top 25%
566 docking score and top 25% W_{QB} values, where optimal enrichment factors (EF) are
567 achieved. **d.** For the same set, distribution of W_{QB} values for the active compounds
568 (red), all decoys (black) and decoys in the top 25% docking score (gray). **e.**
569 Distribution of W_{QB} values of CDK2 actives (red) and decoys (gray) ranked in the
570 top 25% by two independent docking programs (rDock and Glide). **f.** Distribution
571 of W_{QB} values of CDK2 actives (red) and decoys (gray) ranked in the top 25% both
572 by MMPBSA and the rDock docking program.

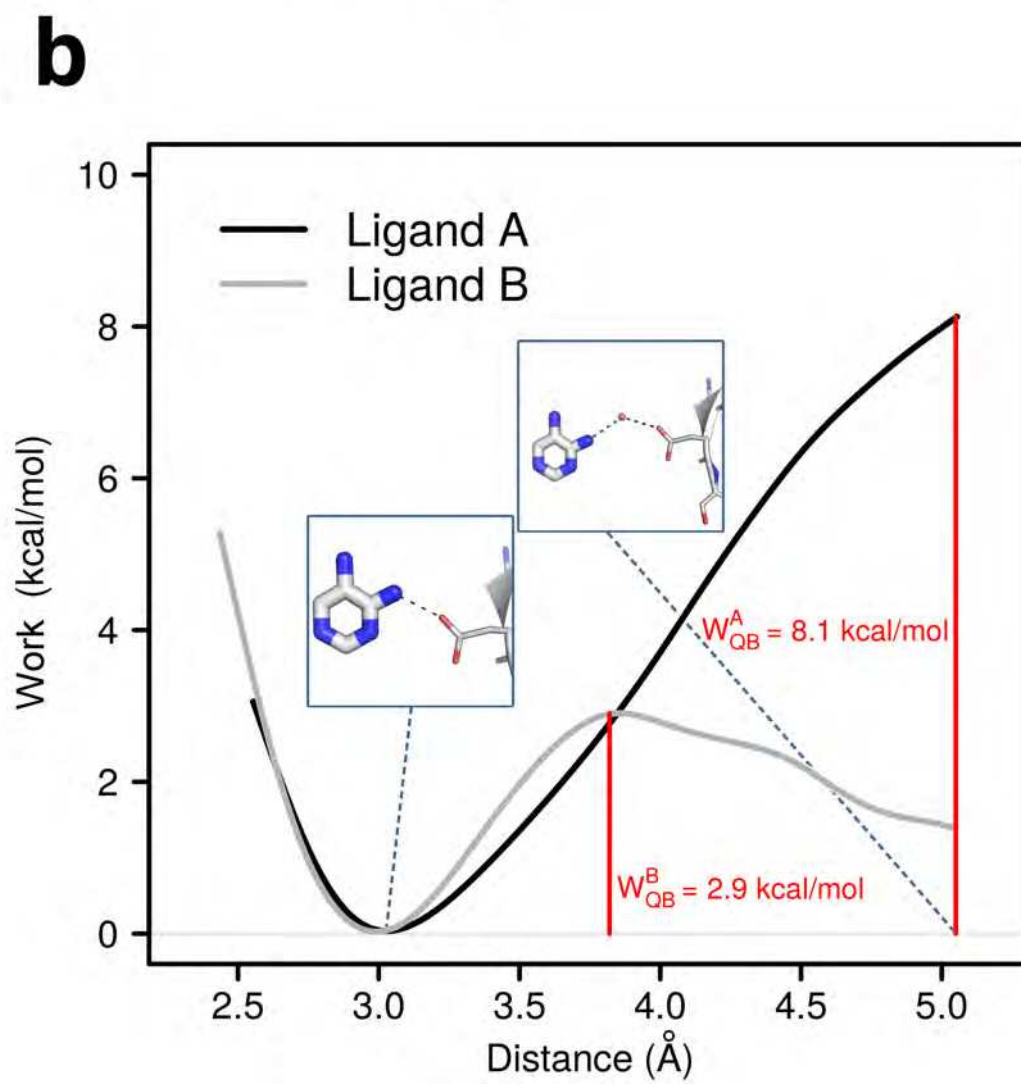
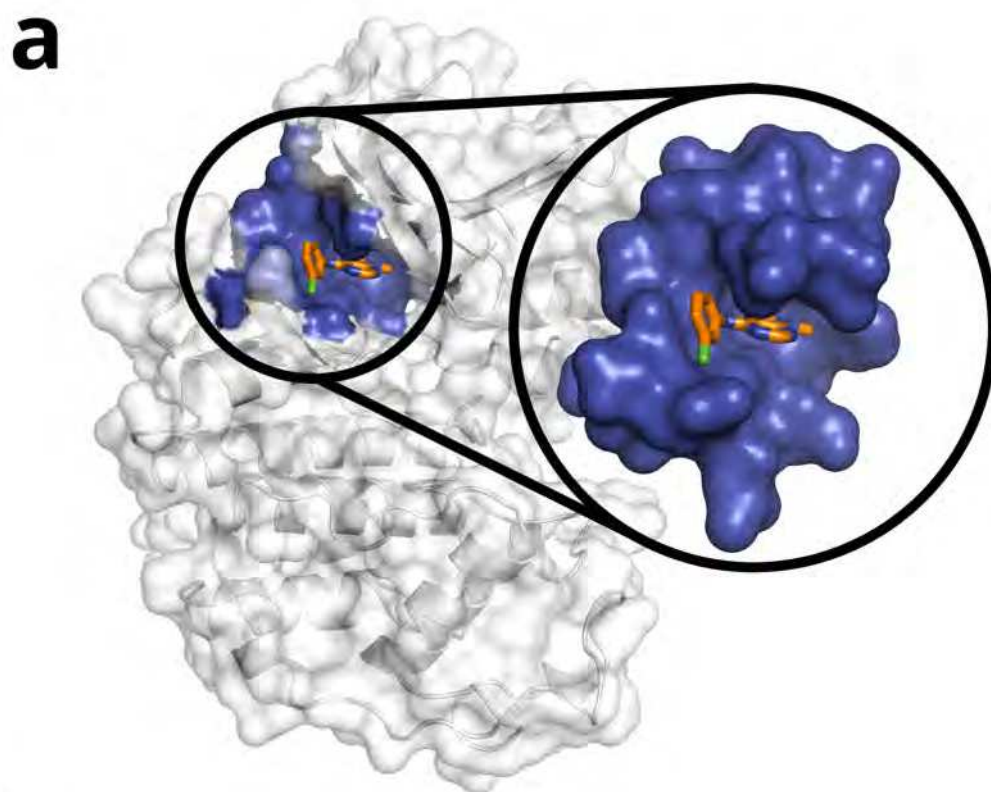
573

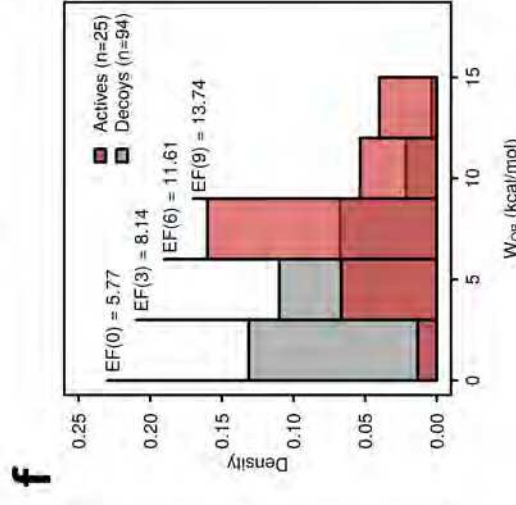
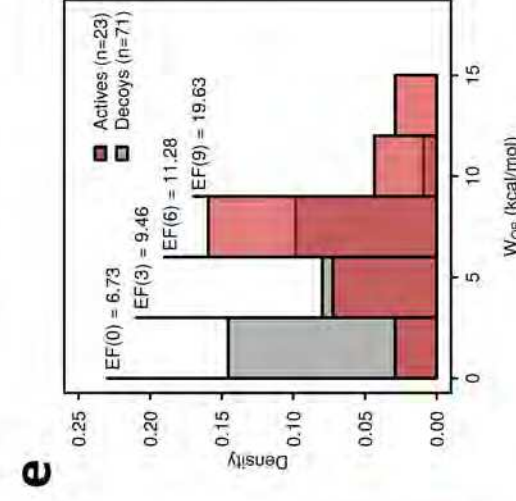
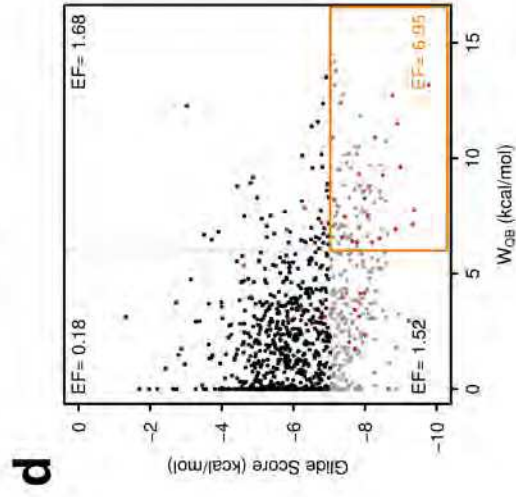
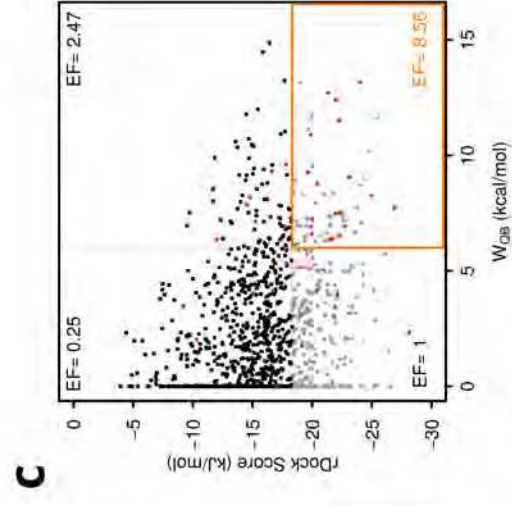
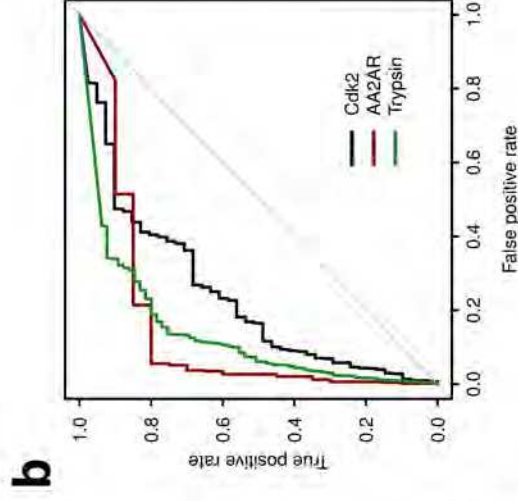
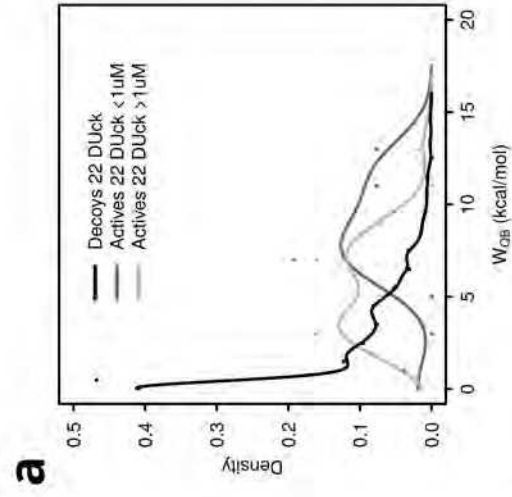
574

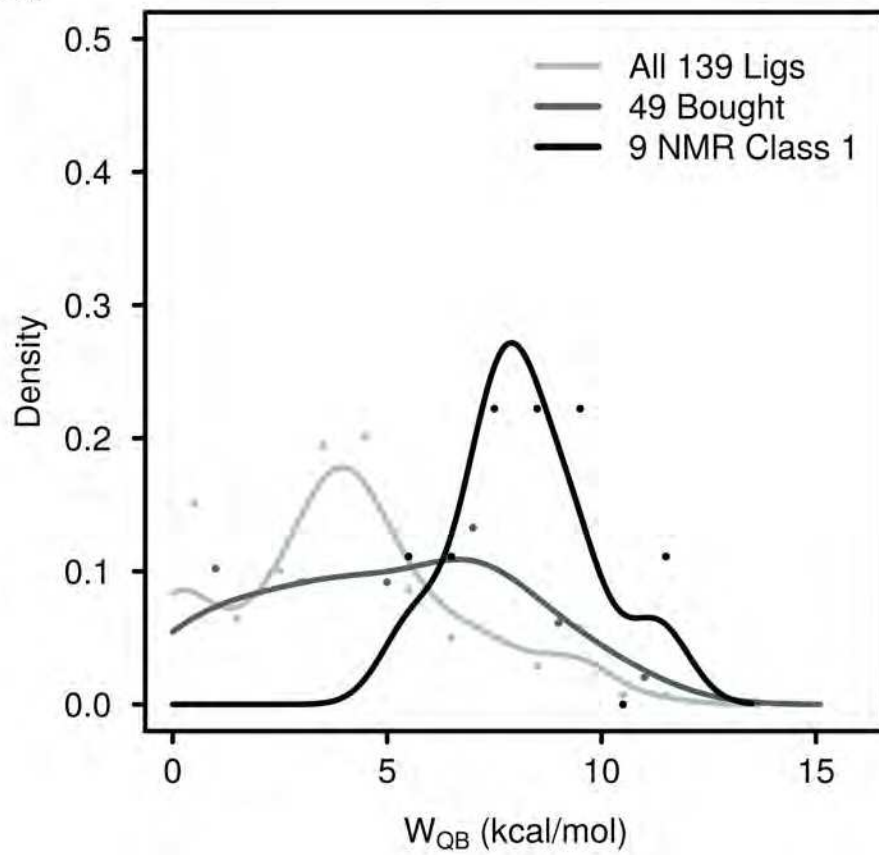
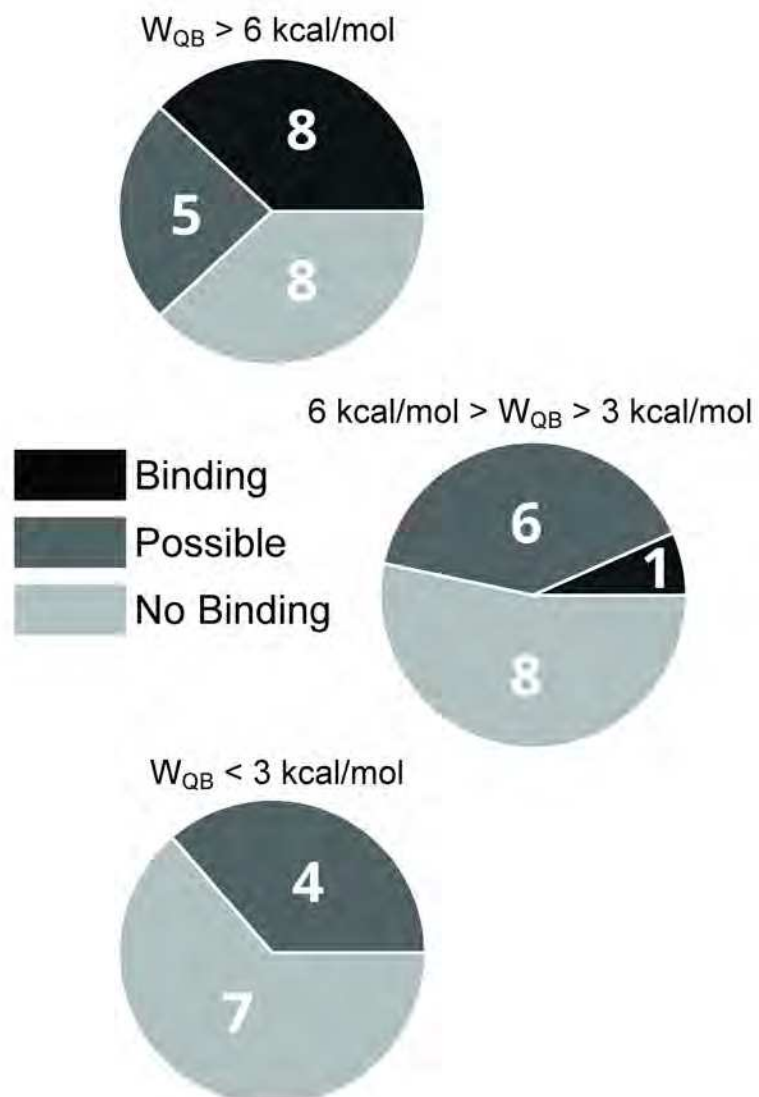
575 **Figure 3.** Additional analyses of the prospective application of DUck in Hsp90. **a.**
576 Distribution of W_{QB} values for 139 top docking scorers (pale gray), 47 compounds
577 within this set that were purchased (dark gray), and the 9 compounds detected as
578 active. **b.** Pie charts showing the hit rates for the set of compounds with high W_{QB}
579 (top), medium W_{QB} (middle) and low W_{QB} (bottom). The area in black corresponds
580 to bona fide hits, dark gray represents compounds that give a positive signal in 1
581 or 2 NMR experiments, pale gray corresponds to inactive compounds. Labels
582 indicate the number of compounds of each class. Chemical structures are shown in
583 Supplementary Figure 12.

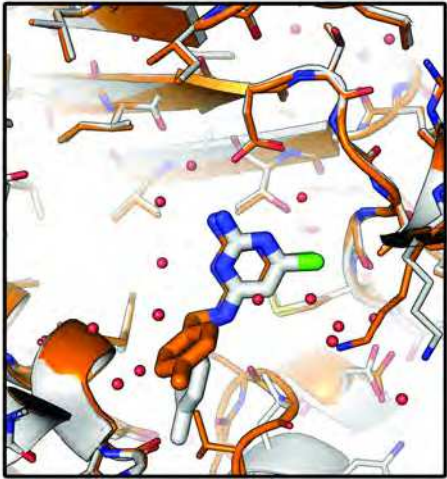
584

585 **Figure 4.** Experimental (grey) and predicted (orange) binding modes of the
586 fragment hits. **a.** Compound **1**, the RMSD of the whole molecule is 2.58 Å due to a
587 conformational change of the protein next to the p-toluene ring. The pyridine ring
588 and bonded atoms, where the key interaction occur, have a RMSD of 0.54 Å **b.**
589 Compound **2** has a RMSD of 0.54 Å **c.** Compound **3** has a RMSD of 1.55 Å, all
590 hydrogen bond interactions are preserved.
591

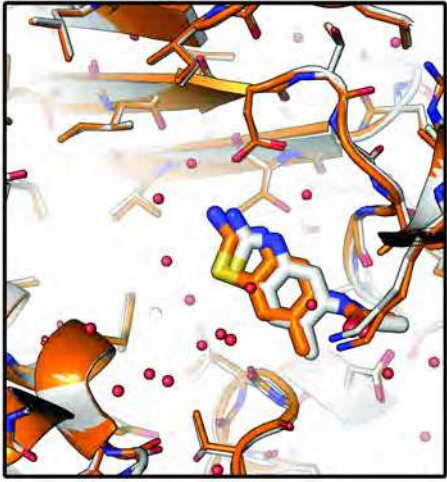




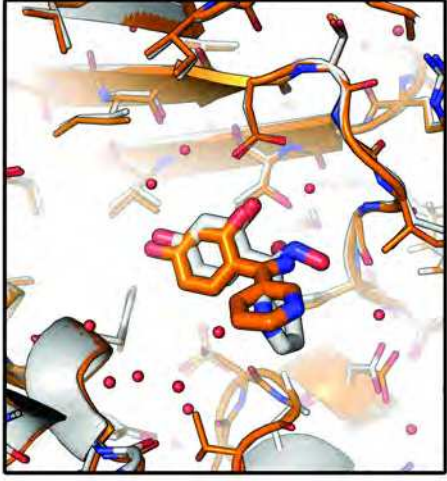
a**b**



a



b



c

Supplementary Information

Dynamic Undocking and the Quasi-Bound State as Tools for Drug Design

Sergio Ruiz-Carmona, Peter Schmidtke, F. Javier Luque, Lisa Baker, Natalia Matassova, Ben
Davis, Stephen Roughley, James Murray, Rod Hubbard, Xavier Barril

INDEX

SUPPLEMENTARY METHODS	3
DATASETS	3
MOLECULAR DOCKING WITH RDOCK	4
MOLECULAR DOCKING WITH GLIDE	4
MMGBSA AND MMPBSA	4
SURFACE PLASMON RESONANCE	5
SUPPLEMENTARY REFERENCES	6
SUPPLEMENTARY FIGURES	7
SUPPLEMENTARY FIGURE 1	7
SUPPLEMENTARY FIGURE 2	7
SUPPLEMENTARY FIGURE 3	8
SUPPLEMENTARY FIGURE 4	9
SUPPLEMENTARY FIGURE 5	9
SUPPLEMENTARY FIGURE 6	10
SUPPLEMENTARY FIGURE 7	11
SUPPLEMENTARY FIGURE 8	12
SUPPLEMENTARY FIGURE 9	13
SUPPLEMENTARY FIGURE 10	14
SUPPLEMENTARY FIGURE 11	15
SUPPLEMENTARY FIGURE 12	16
SUPPLEMENTARY FIGURE 13	19
SUPPLEMENTARY FIGURE 14	20
SUPPLEMENTARY FIGURE 15	21
SUPPLEMENTARY FIGURE 16	22
SUPPLEMENTARY FIGURE 17	23
SUPPLEMENTARY FIGURE 18	24
SUPPLEMENTARY FIGURE 19	24
SUPPLEMENTARY FIGURE 20	25
SUPPLEMENTARY FIGURE 21	26
SUPPLEMENTARY FIGURE 22	27
SUPPLEMENTARY FIGURE 23	28
SUPPLEMENTARY FIGURE 24	29
SUPPLEMENTARY FIGURE 25	29
SUPPLEMENTARY FIGURE 26	30
SUPPLEMENTARY TABLES	31
SUPPLEMENTARY TABLE 1	31
SUPPLEMENTARY TABLE 2	32
SUPPLEMENTARY TABLE 3	32
SUPPLEMENTARY TABLE 4	33
SUPPLEMENTARY TABLE 5	34
SUPPLEMENTARY TABLE 6	35

SUPPLEMENTARY METHODS

Datasets

When possible, datasets were geared towards fragment-sized ligands because they present more scaffold diversity, make fewer peripheral interactions that could mask the main interactions and because Fragment-Based Drug Discovery (FBDD) approaches are increasingly important as hit identification strategy.^{1,2} For CDK2, all ligands with molecular weight below 300 Da and known binding affinity (IC₅₀) were extracted from the PDB.³ To increase the diversity of the dataset, all ligands were clustered at 75% similarity using the MACCS fingerprints as implemented in MOE (Chemical Computing Group Inc., 2015) and only the centroids were used to define the active set. The composition of the dataset is described in Supplementary Table 5. It should be noted that this is a noisy dataset because data sources are very heterogeneous and IC₅₀ values have an indirect relationship with dissociation constants.⁴ As such, it should only be used to detect trends. In order to assess the significance of the correlation, we have also investigated the correlation between IC₅₀ and molecular weight (Supplementary Figure 20). For retrospective VS experiments, a pool of 30 decoys per active fragment was obtained with the DUD-E decoy generator,⁵ which puts together a set of putatively inactive molecules with physicochemical properties very similar to active ones. For BRD4, as it was designed to study the correlation between experimental binding affinity and W_{QB} , only the ligands with known binding mode and measured IC₅₀ or K_D were considered (relationship with molecular weight reported in Supplementary Figure 21). The crystal structure of each ligand-protein complex was obtained from PDB and used as input for subsequent calculations. The composition of the dataset is described in Supplementary Table 6. In the case of AA2AR, as there are few structures in the PDB, the active fragments were taken from the DUD-E benchmark set.⁵ The rest of the procedure is the same as described for CDK2. For Trypsin, we found that few ligands have a low molecular weight so we did not filter by size. Instead, a random subset of 2000 actives and decoys was selected from DUD-E. In the case of Hsp90, all candidate molecules originate from a unified collection generated in house from the commercial libraries of five preferred vendors (Specs, Enamine, Life Chemicals, Princeton Biomoleculars and Asinex). In this case we set an upper limit of 250 Da, obtaining 280000 candidate fragments. All ligands were

prepared for docking using Schrödinger's Ligprep⁶ with the following options different than default: neutralize and ionize at pH 7 with a threshold of +/- 1 with a maximum of 6 tautomers and 8 stereoisomers generated.

Molecular Docking with rDock

For CDK2, AA2AR and Trypsin, the 3D structure used to define the receptor was obtained from the DUD-E benchmark set.⁵ MOE⁷ was used to generate mol2 files that can be read by rDock, our docking engine.⁸ For Hsp90, we use the same cavity definition and docking protocol described previously.⁸ In all systems, pharmacophoric restraints were used to ensure that the key interaction point was matched by every molecule in the dataset, as defined in Supplementary Table 3. rDock was run with the default parameters for standard docking. 50 individual docking processes were executed per ligand, thus ensuring that the lowest-energy binding mode is identified. The best-scoring solution is accepted as the putative binding mode. Ligands that do not fulfill the pharmacophore are identified by the restraint penalty and eliminated from the dataset (i.e. not considered in the ROC curves or any other analysis).

Molecular Docking with Glide

In order to demonstrate that our methodology provides an advantage regardless of the docking program used, we also run CDK2, AA2R and trypsin systems with Glide.⁹ The generation of the cavity with Glide was performed using coordinates defined as in rDock docking and default parameters. Pharmacophoric restraints were defined to force all ligands to make a hydrogen bond as defined in Supplementary Table 3. Glide docking was run with default parameters and with pharmacophoric restraints (Supplementary Figures 22, 23 and 24). The best docking pose for each ligand was selected and used as input for DUck.

MMGBSA and MMPBSA

MMGBSA and MMPBSA calculations using AMBER12 software were also performed and compared against the rest of methods. Each ligand was simulated for 5 ns with the full size receptor of CDK2 using the same MD configuration defined in the section above (Supplementary Figures 24 and 25). For each

simulation, a total of 25 snapshots separated by 200 ps were used and the free energies were averaged over the ensemble of conformations. All the calculations were performed with default parameters with the exception of the following: the GB model used is one of the developed by Onufriev et al.¹⁰ (igb=2) and the atomic radii are set up according to the topology (radiopt=0).

Surface plasmon resonance

Surface plasmon resonance (SPR) experiments have been done mainly as described before.^{11,12} All measurements were performed on a Biacore T200 instrument (Biacore GE Healthcare) at 20°C on Series S NTA chips. 25 mM HEPES pH7.4, 175 mM NaCl, 0.01% P-20, 0.025mM EDTA and 1% DMSO was used as a running buffer. HSP90 protein was produced as described previously. Chip surface was generated with multi-His-tagged Hsp90 protein as in reference.¹¹ The sensor surface was regenerated by 0.35 M EDTA and 45% DMSO with additional 60 sec injections of 0.1 mg/mL trypsin and 0.5 M imidazole. In some experiments, the protein was further stabilized on NTA surface by covalent amine coupling as advised by manufacturer. Screening of fragments was conducted in dose response titrations of nine two-fold diluted experimental points with the top concentration of 500 μ M. Each fragment has been tested at least three times. Data processing was performed using BIAevaluation 2.1 (Biacore GE Healthcare Bio-SciencesCorp) or Scrubber2 (BioLogic) software. Sensorgrams were double referenced prior to global fitting of the concentration series to a Steady State Affinity model. Representative sensorgrams are shown in Supplementary Figure 26.

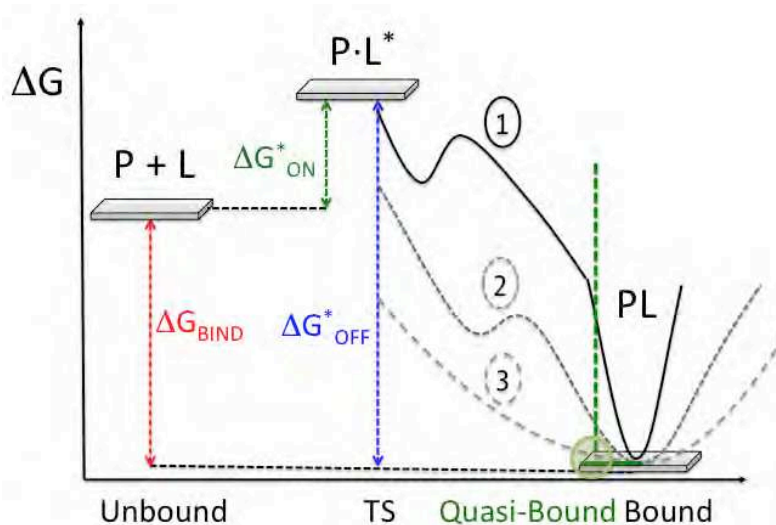
SUPPLEMENTARY REFERENCES

1. Joseph-McCarthy, D., Campbell, A. J., Kern, G. & Moustakas, D. Fragment-based lead discovery and design. *J. Chem. Inf. Model.* **54**, 693–704 (2014).
2. Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **6**, 211–9 (2007).
3. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
4. Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discov.* **14**, 95–110 (2015).
5. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
6. Schrödinger. LigPrep, version 2.9. *Schrödinger, LLC* (2014).
7. Chemical Computing Group Inc. Molecular Operating Environment (MOE), 2014.09. (2015).
8. Ruiz-Carmona, S. *et al.* rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **10**, e1003571 (2014).
9. Friesner, R. a *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–49 (2004).
10. Onufriev, A., Bashford, D. & Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins Struct. Funct. Genet.* **55**, 383–394 (2004).
11. Meiby, E. *et al.* Fragment screening by weak affinity chromatography: comparison with established techniques for screening against HSP90. *Anal. Chem.* **85**, 6756–66 (2013).
12. Murray, J. B., Roughley, S. D., Matassova, N. & Brough, P. a. Off-rate screening (ORS) by surface plasmon resonance. An efficient method to kinetically sample hit to lead chemical space from unpurified reaction products. *J. Med. Chem.* **57**, 2845–2850 (2014).
13. Anderson, D. R. *et al.* Benzothiophene inhibitors of MK2. Part 1: structure-activity relationships, assessments of selectivity and cellular potency. *Bioorg. Med. Chem. Lett.* **19**, 4878–81 (2009).
14. O'Boyle, N. M. *et al.* Open Babel: An Open chemical toolbox. *J. Cheminform.* **3**, (2011).

SUPPLEMENTARY FIGURES

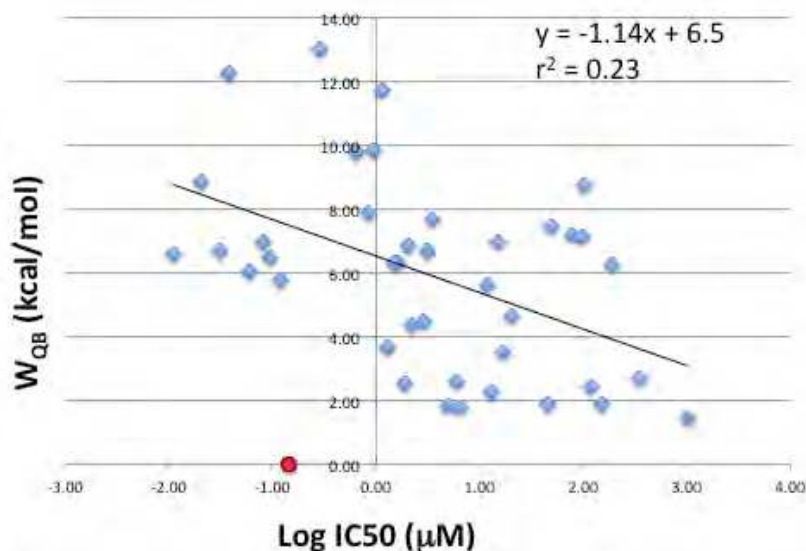
Supplementary Figure 1

Graphical representation of the quasi-bound state in relation to the dissociation process. The macroscopic constants describing the behavior of a non-covalent complex are determined by the relative free energies of three states (bound, transition state and unbound). States in-between are theoretically irrelevant, so molecules 1, 2 and 3 would have the same kinetic and thermodynamic constants. The Quasi-bound state is merely designed to probe the slope around the bound state, obtaining an approximation to the structural stability of the binding mode. We find that true ligands are more likely to have a profile like 1, whereas many decoys have profiles similar to 2 or 3.



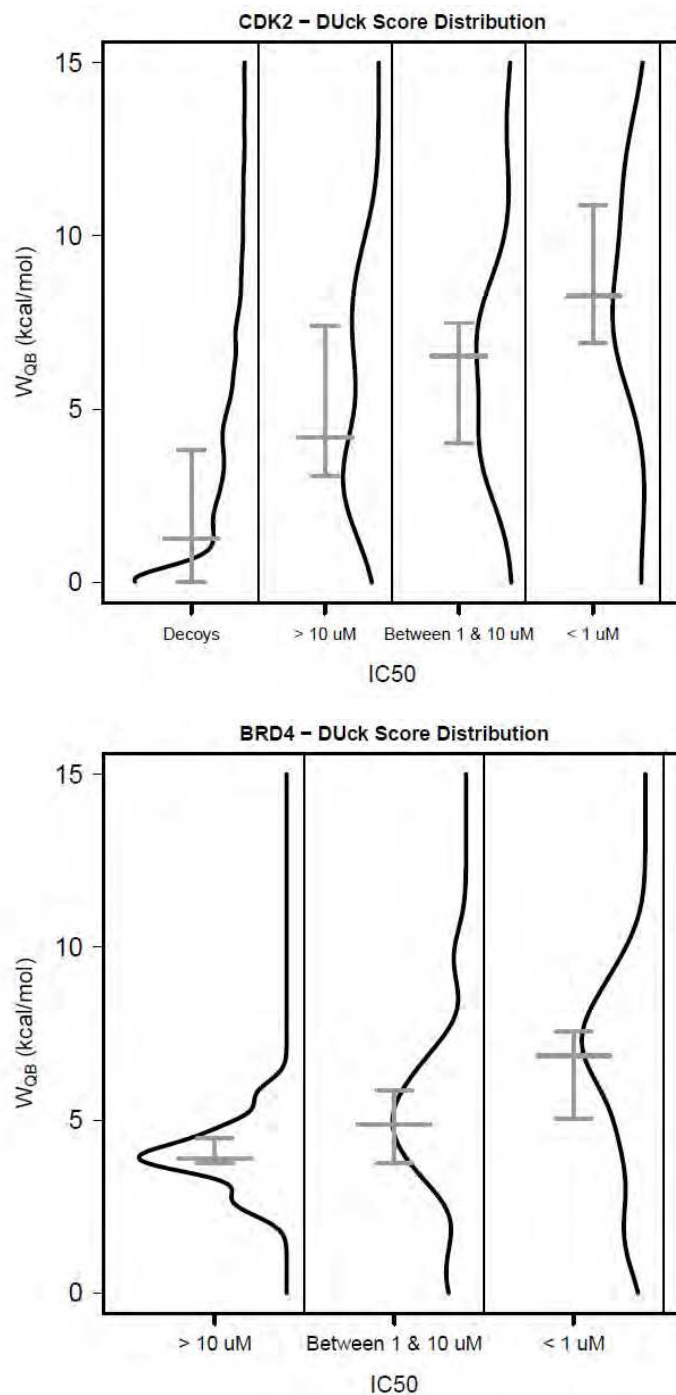
Supplementary Figure 2

W_{QB} values vs. experimentally determined activities (expressed as $\text{Log}(IC_{50})$), for a set of 41 Fragment-like CDK2 ligands taken from the PDB. Ligand 3FZ1 is shown in red and not included in the correlation. As shown below, this ligand does not fulfill the condition of using the hinge region as attachment point.



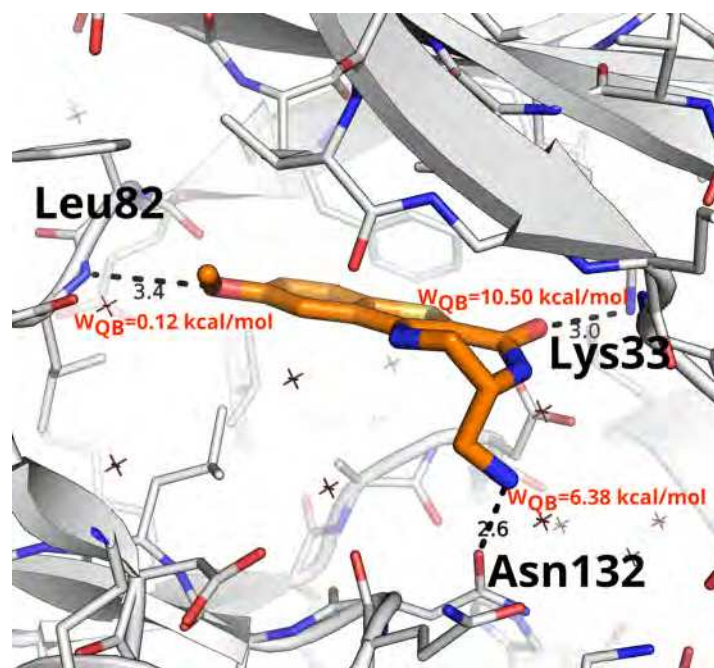
Supplementary Figure 3

Distribution of W_{QB} values as a function of binding affinity (IC_{50}), for the CDK2 (top) and BRD4 set (bottom). Compounds with the same binding affinity present a wide distribution of W_{QB} values, but there is a tendency towards higher values for more potent compounds. Most notably, very low W_{QB} values are rare for potent ligands.

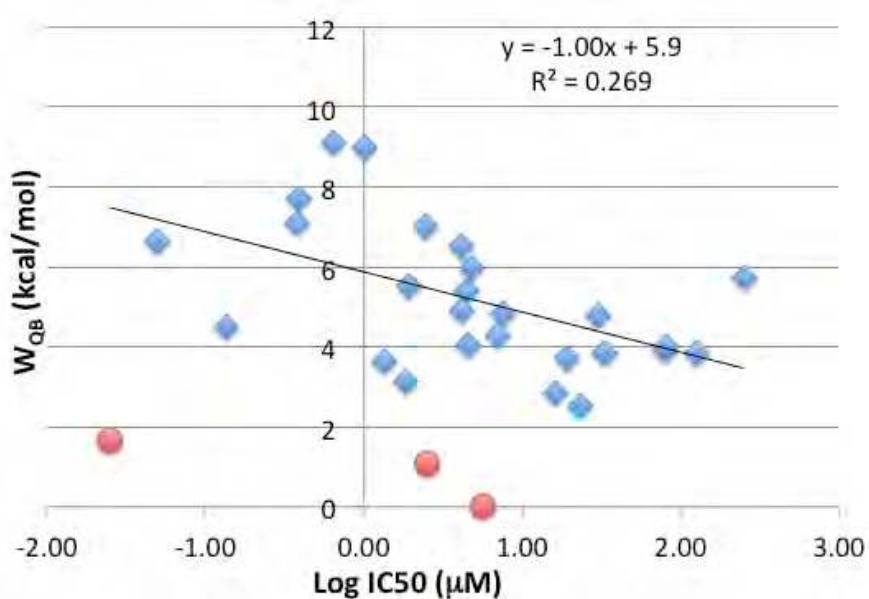


Supplementary Figure 4

Binding mode of ligand in PDB structure 3FZ1. This ligand is unusual because its interaction with the hinge region is labile. Structural and SAR data confirms that this interaction is not important for potency.¹³ Instead, this ligand forms two charge-reinforced hydrogen bonds with N ζ of Lys33 and O δ 1 of Asn132, from which it draws structural stability ($W_{QB} = 10.50$ kcal/mol; Supplementary Figure 18). Note that the IC₅₀ reported in the PDB for this compound is wrong. The correct value is 146 nM.¹³

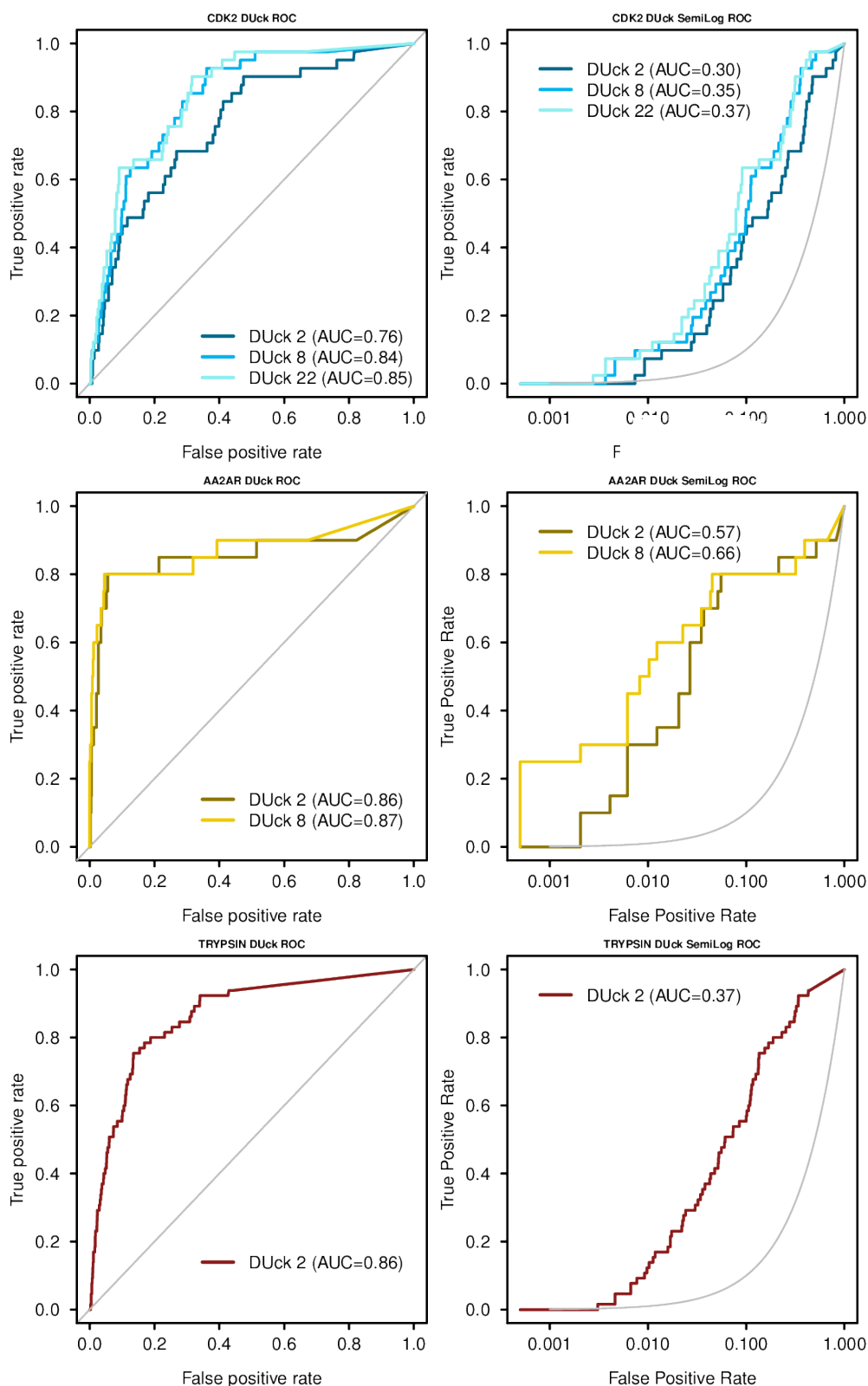
**Supplementary Figure 5**

W_{QB} values vs. experimentally determined activities (expressed as $\text{Log}(\text{IC}_{50})$), for a set of 30 BRD4 ligands taken from the PDB. The points in red have not been included in the correlation. They correspond to three kinase inhibitors that bind to BRD4 as an unintended secondary target and present extremely low resistance to breaking the interaction with N δ 2 of Asn120 (PDB codes 4074, 4077 & 407E).



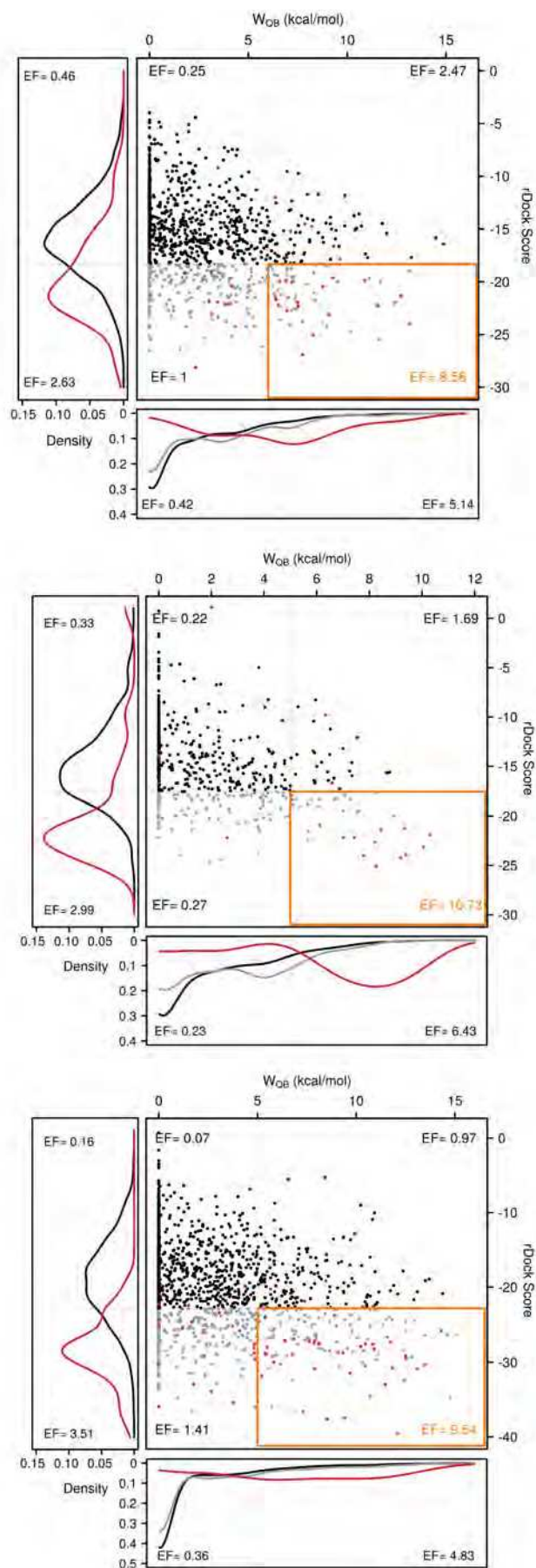
Supplementary Figure 6

ROC curves (left) and semilog-ROC curves (right) of the retrospective virtual screening experiments on CDK2 (top), AA2R (middle) and Trypsin (bottom). The grey line indicates the baseline (random selection). For CDK2, the results corresponding to 2, 8 and 22 DUck runs are reported. For AA2R, the results corresponding to 2, and 8 DUck runs are reported. For Trypsin, only 2 DUck runs were executed. AUC values are inset in the plots.



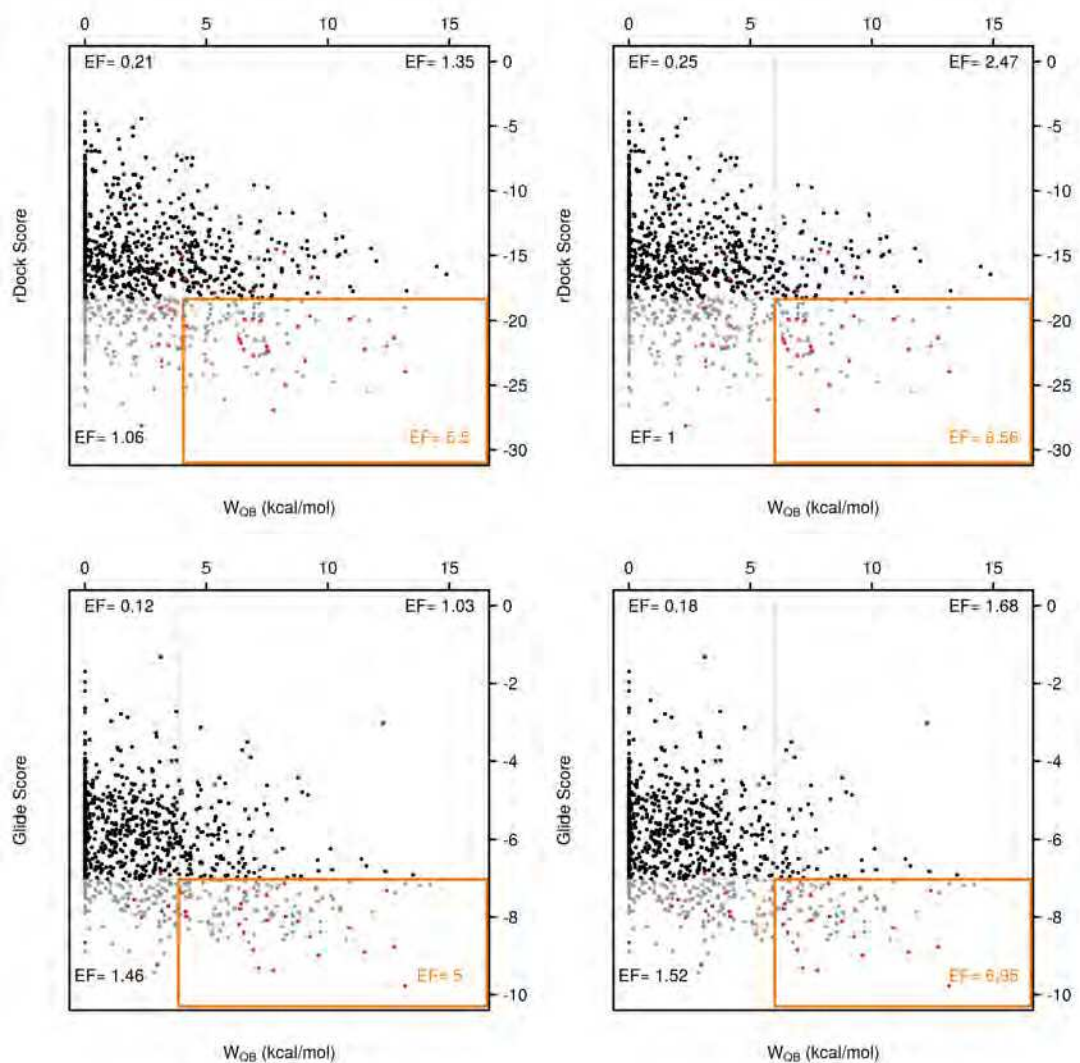
Supplementary Figure 7

Docking (rDock) score vs. W_{QB} values for active (red) and inactive compounds (black or gray) in the retrospective virtual screening datasets for CDK2 (top), AA2AR (middle) and Trypsin (bottom). The side panels show the distribution of active (red) and inactive (black) compounds for each individual method (docking to the left, DUCK at the bottom). Gray points (central panel) and gray line (inferior panel) represent the decoys with a docking score within the top 25%. The orange square highlights the area corresponding to top 25% docking score and top 25% W_{QB} values, where optimal enrichment factors (EF) are achieved.



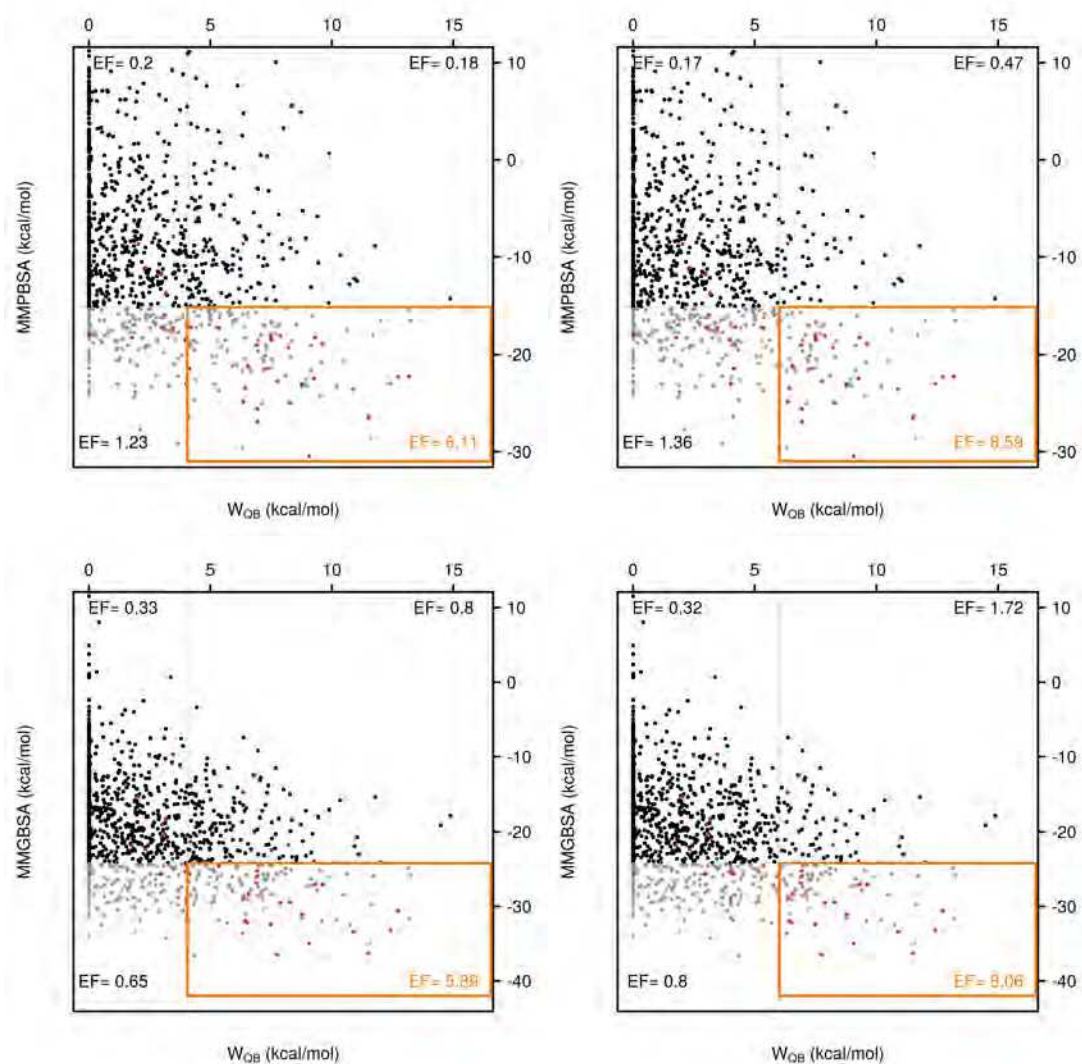
Supplementary Figure 8

Docking score vs. W_{QB} obtained for two different programs on the CDK2 test set. Each molecule was docked with rDock (top) or Glide (bottom) and the binding mode generated by each program was used as starting geometry for DUck simulations. In both cases, docking scores are orthogonal to W_{QB} and a high proportion of good scorers have very low W_{QB} values. The intersection between methods defines a subset highly enriched in active molecules. Two intersecting levels are presented per program: top25% docking+ top 25% DUck (left); and top25% docking + top 12% DUck (right).



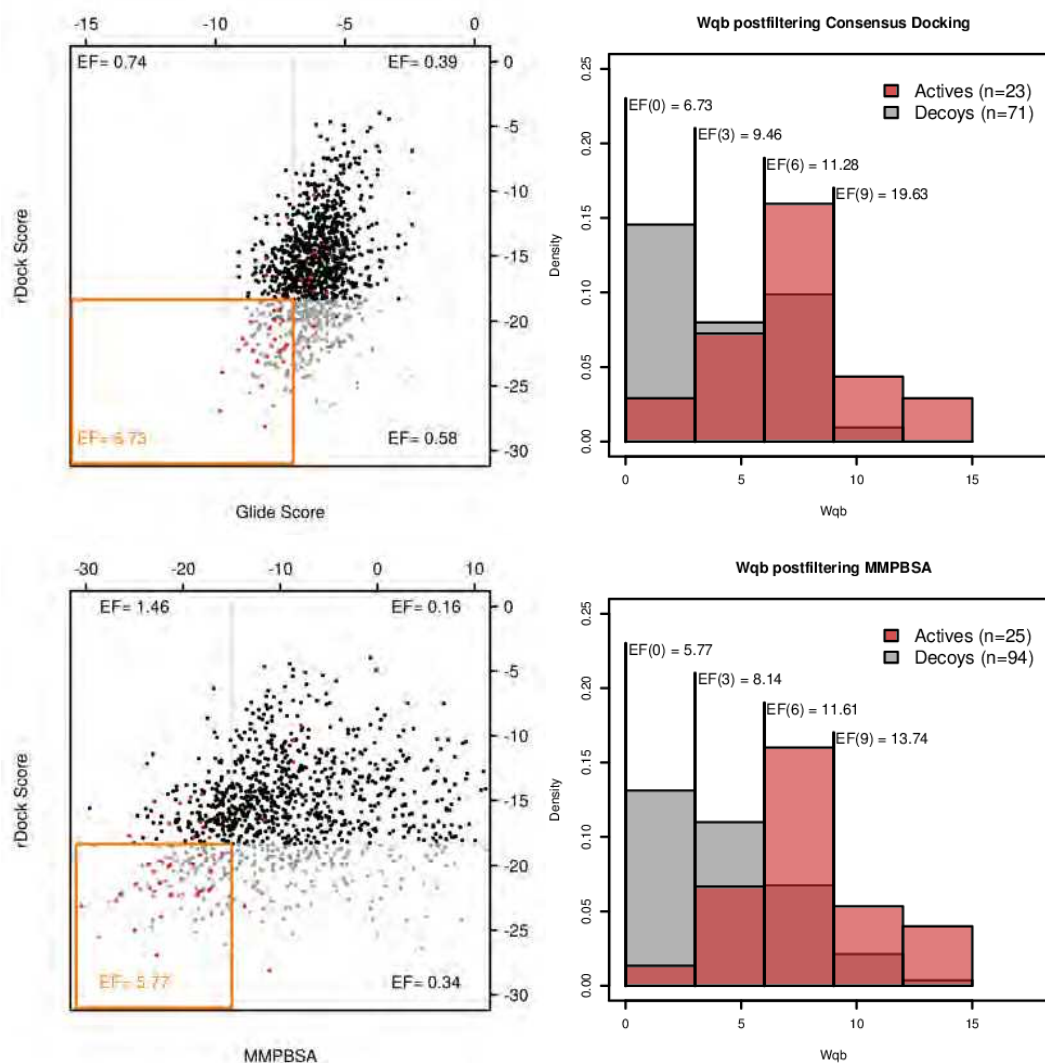
Supplementary Figure 9

MMPBSA and MMGBSA-calculated ΔG_{bind} vs. W_{QB} on the CDK2 test set. The rDock-generated binding mode was used as starting point for molecular dynamics simulations, which were then processed to obtain MMPBSA and MMGBSA binding free energies. In both cases, the calculated ΔG_{bind} values are orthogonal to W_{QB} and a high proportion of good MM(PB/GB)SA scorers have very low W_{QB} values. The intersection between methods defines a subset highly enriched in active molecules. Two intersecting levels are presented per method: top25% MM(PB/GB)SA + top 25% DUck (left); and top25% MM(PB/GB)SA + top 12% DUck (right).



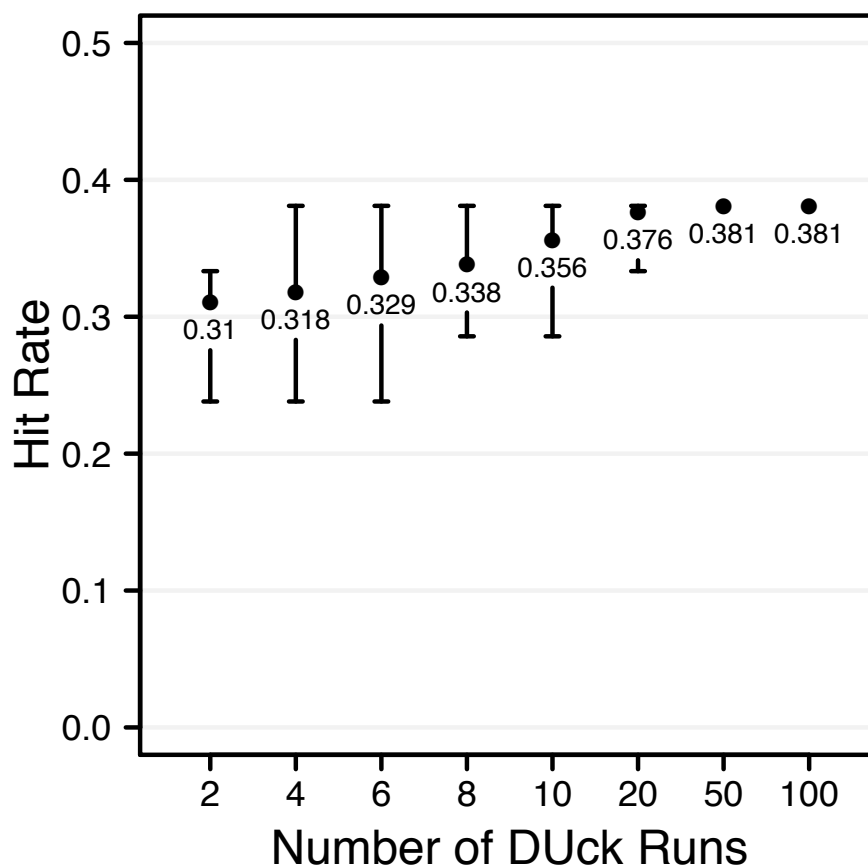
Supplementary Figure 10

Filtering by W_{QB} increases performance even after consensus scoring (CDK2 test set). The left panels show a scatter plot of rDock score vs. Glide score (top) and rDock score vs. MMPBSA-calculated ΔG_{bind} (bottom). Molecules ranked in the top 25% by both methods (highlighted area) are then binned according to their W_{QB} (right panels, also shown in the main text). Filtering by W_{QB} would increase the enrichment factor in a cut-off dependent manner.



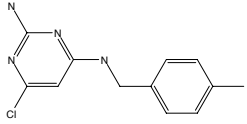
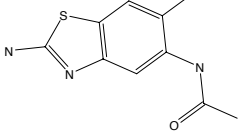
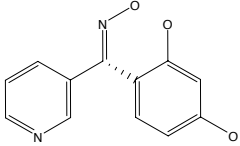
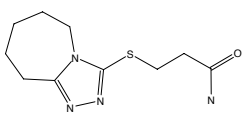
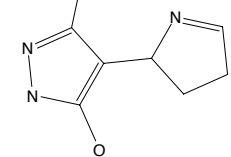
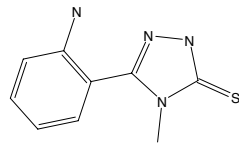
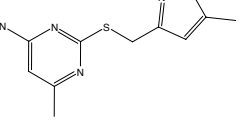
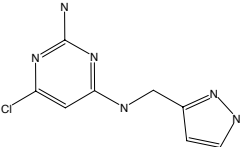
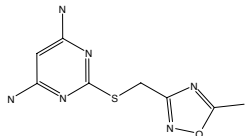
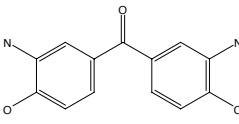
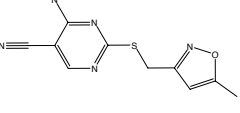
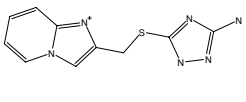
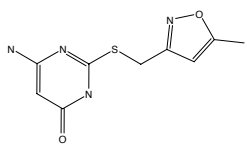
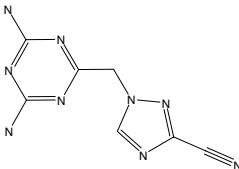
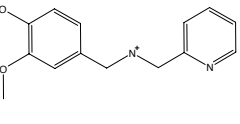
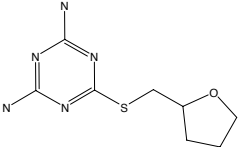
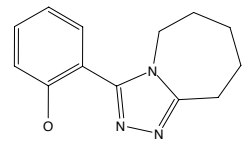
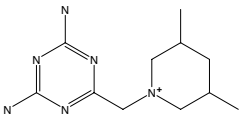
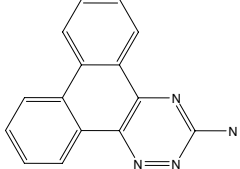
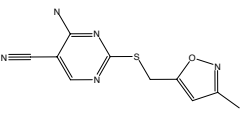
Supplementary Figure 11

Percentage of active molecules in the top 21 (out of 47 compounds tested) as a function of the number of DUck runs. At the screening stage we carried out 100 DUck simulations per ligand, obtaining a hit rate of 38%. Retrospectively, we took 50 random combinations of $N=\{2,4,6,8,10,20,50\}$ DUck runs and calculated the hit rates that would have been obtained. Averages are represented as filled circles and labeled with their actual values. The bars span from the maximum to the minimum values.

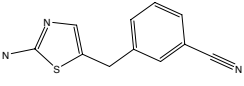
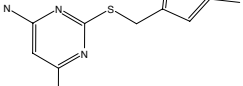
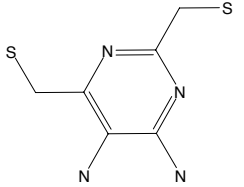
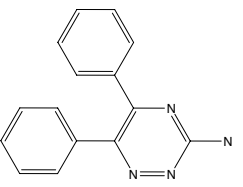
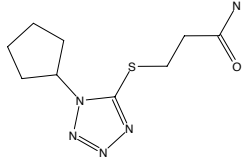
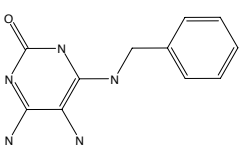
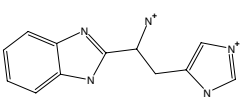
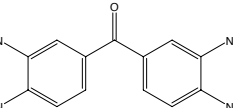
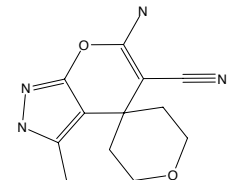
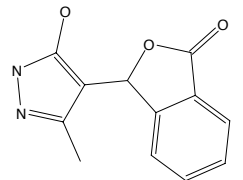
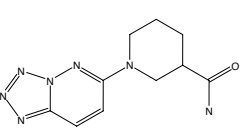
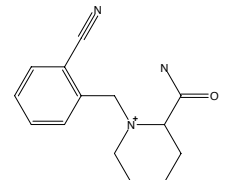
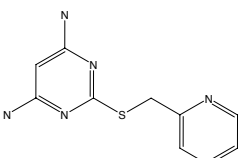
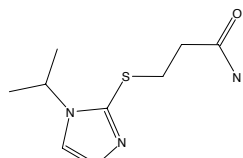
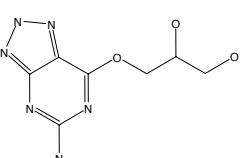
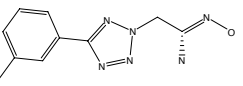
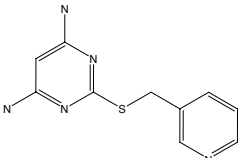
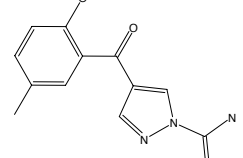
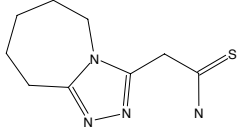
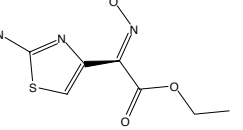


Supplementary Figure 12

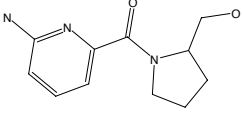
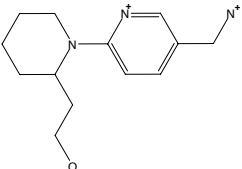
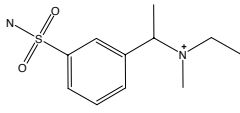
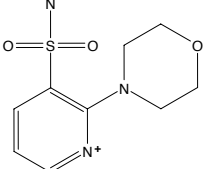
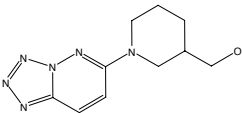
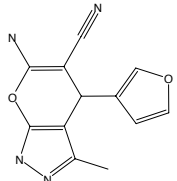
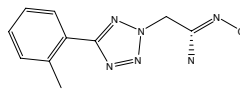
Chemical structure of the tested compounds. Duck Class refers to strong, medium and weak binders (1, 2 & 3, respectively). NMR Class 1 are true binders. The rest are considered inactive. The real numbers correspond to rDock score (left) and W_{QB} (right).

<p>Duck Class: 1</p>  <p>NMR Class: 1 -24.9700 9.1000 ID: 1</p>	<p>Duck Class: 1</p>  <p>NMR Class: 1 -25.0300 8.2000 ID: 2</p>	<p>Duck Class: 1</p>  <p>NMR Class: 1 -26.6200 11.3000 ID: 3</p>	<p>Duck Class: 1</p>  <p>NMR Class: 1 -26.4500 7.4000 ID: 4</p>
<p>Duck Class: 1</p>  <p>NMR Class: 1 -23.7700 8.2000 ID: 5</p>	<p>Duck Class: 1</p>  <p>NMR Class: 1 -23.2600 9.5000 ID: 6</p>	<p>Duck Class: 1</p>  <p>NMR Class: 1 -25.4500 7.8000 ID: 7</p>	<p>Duck Class: 1</p>  <p>NMR Class: 1 -25.3500 7.0000 ID: 8</p>
<p>Duck Class: 1</p>  <p>NMR Class: 2 -28.0400 7.3000 ID: 9</p>	<p>Duck Class: 1</p>  <p>NMR Class: 2 -27.1200 6.4000 ID: 10</p>	<p>Duck Class: 1</p>  <p>NMR Class: 2 -26.1400 9.8000 ID: 11</p>	<p>Duck Class: 1</p>  <p>NMR Class: 3 -25.4000 7.0000 ID: 12</p>
<p>Duck Class: 1</p>  <p>NMR Class: 3 -25.0100 6.4000 ID: 13</p>	<p>Duck Class: 1</p>  <p>NMR Class: nb -26.0000 6.5000 ID: 14</p>	<p>Duck Class: 1</p>  <p>NMR Class: nb -25.9100 6.7000 ID: 15</p>	<p>Duck Class: 1</p>  <p>NMR Class: nb -24.7300 6.5000 ID: 16</p>
<p>Duck Class: 1</p>  <p>NMR Class: nb -24.4600 10.3000 ID: 17</p>	<p>Duck Class: 1</p>  <p>NMR Class: nb -28.4400 7.2000 ID: 18</p>	<p>Duck Class: 1</p>  <p>NMR Class: nb -25.9200 9.2000 ID: 19</p>	<p>Duck Class: 1</p>  <p>NMR Class: nc -26.3400 7.3000 ID: 20</p>

Supplementary Figure 12 (cont)

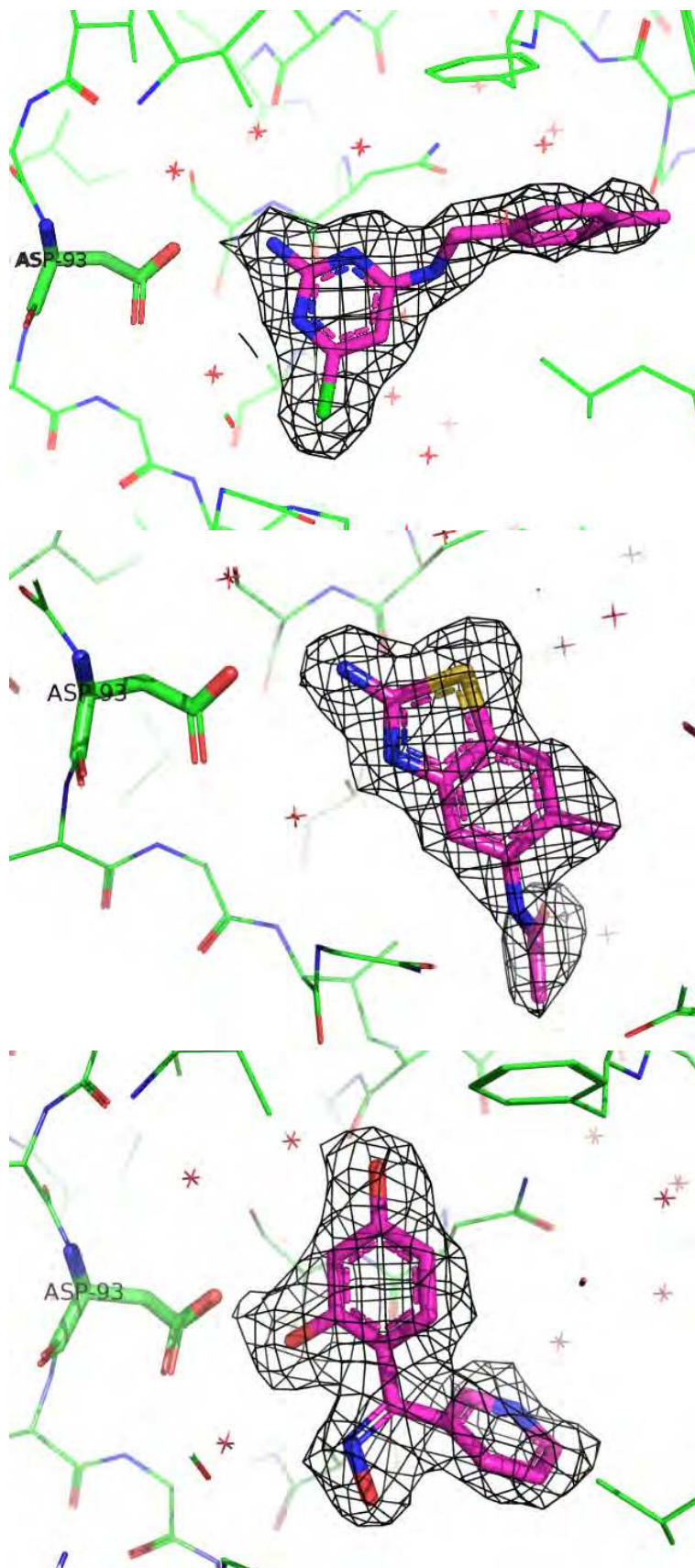
<p>DUck Class: 1</p>  <p>NMR Class: r -24.0100 8.0000 ID: 21</p>	<p>DUck Class: 2</p>  <p>NMR Class: 1 -28.2700 5.6000 ID: 22</p>	<p>DUck Class: 2</p>  <p>NMR Class: 2 -26.8600 4.2000 ID: 23</p>	<p>DUck Class: 2</p>  <p>NMR Class: 2 -25.3300 3.6000 ID: 24</p>
<p>DUck Class: 2</p>  <p>NMR Class: 2 -25.3200 3.9000 ID: 25</p>	<p>DUck Class: 2</p>  <p>NMR Class: 3 -25.0900 5.5000 ID: 26</p>	<p>DUck Class: 2</p>  <p>NMR Class: 3 -27.3400 3.5000 ID: 27</p>	<p>DUck Class: 2</p>  <p>NMR Class: 3 -27.5500 3.2000 ID: 28</p>
<p>DUck Class: 2</p>  <p>NMR Class: nb -25.4600 4.7000 ID: 29</p>	<p>DUck Class: 2</p>  <p>NMR Class: nb -25.0900 5.5000 ID: 30</p>	<p>DUck Class: 2</p>  <p>NMR Class: nb -26.7600 4.4000 ID: 31</p>	<p>DUck Class: 2</p>  <p>NMR Class: nb -26.4700 3.1000 ID: 32</p>
<p>DUck Class: 2</p>  <p>NMR Class: nb -25.1000 4.9000 ID: 33</p>	<p>DUck Class: 2</p>  <p>NMR Class: nb -24.8300 4.8000 ID: 34</p>	<p>DUck Class: 2</p>  <p>NMR Class: nb -24.3200 4.4000 ID: 35</p>	<p>DUck Class: 2</p>  <p>NMR Class: nb -26.8200 3.0000 ID: 36</p>
<p>DUck Class: 3</p>  <p>NMR Class: 2 -27.9100 0.8000 ID: 37</p>	<p>DUck Class: 3</p>  <p>NMR Class: 2 -26.2500 1.4000 ID: 38</p>	<p>DUck Class: 3</p>  <p>NMR Class: 2 -25.4700 0.5000 ID: 39</p>	<p>DUck Class: 3</p>  <p>NMR Class: 3 -25.3700 1.0000 ID: 40</p>

Supplementary Figure 12 (cont)

<p>DUck Class: 3</p>  <p>NMR Class: nb -26.4800 2.3000 ID: 41</p>	<p>DUck Class: 3</p>  <p>NMR Class: nb -25.7700 2.5000 ID: 42</p>	<p>DUck Class: 3</p>  <p>NMR Class: nb -24.9900 0.0000 ID: 43</p>	<p>DUck Class: 3</p>  <p>NMR Class: nb -24.9700 0.0000 ID: 44</p>
<p>DUck Class: 3</p>  <p>NMR Class: nb -24.9100 0.0000 ID: 45</p>	<p>DUck Class: 3</p>  <p>NMR Class: nb -23.2300 2.0000 ID: 46</p>	<p>DUck Class: 3</p>  <p>NMR Class: nb -26.1100 1.7000 ID: 47</p>	

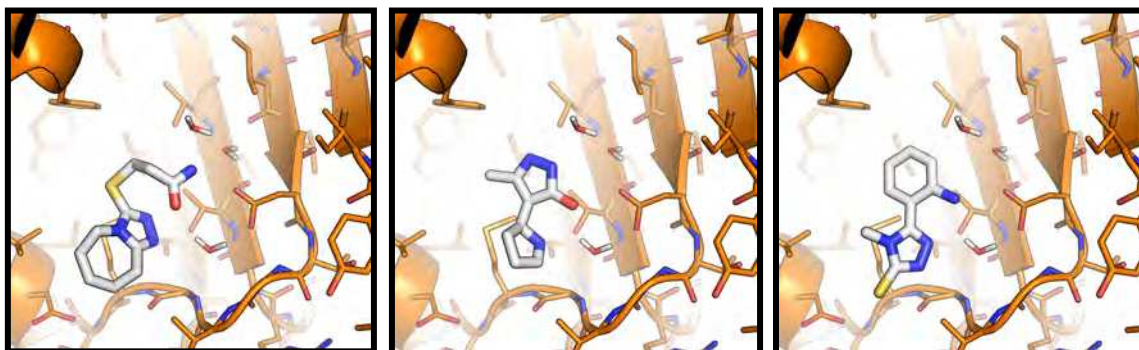
Supplementary Figure 13

Crystal structure of Hsp90 in complex with compounds **1** (top), **2** (middle) and **3** (bottom). The 2fofc electron density maps are displayed at the 1.0 Sigma level (Carve = 1.7).



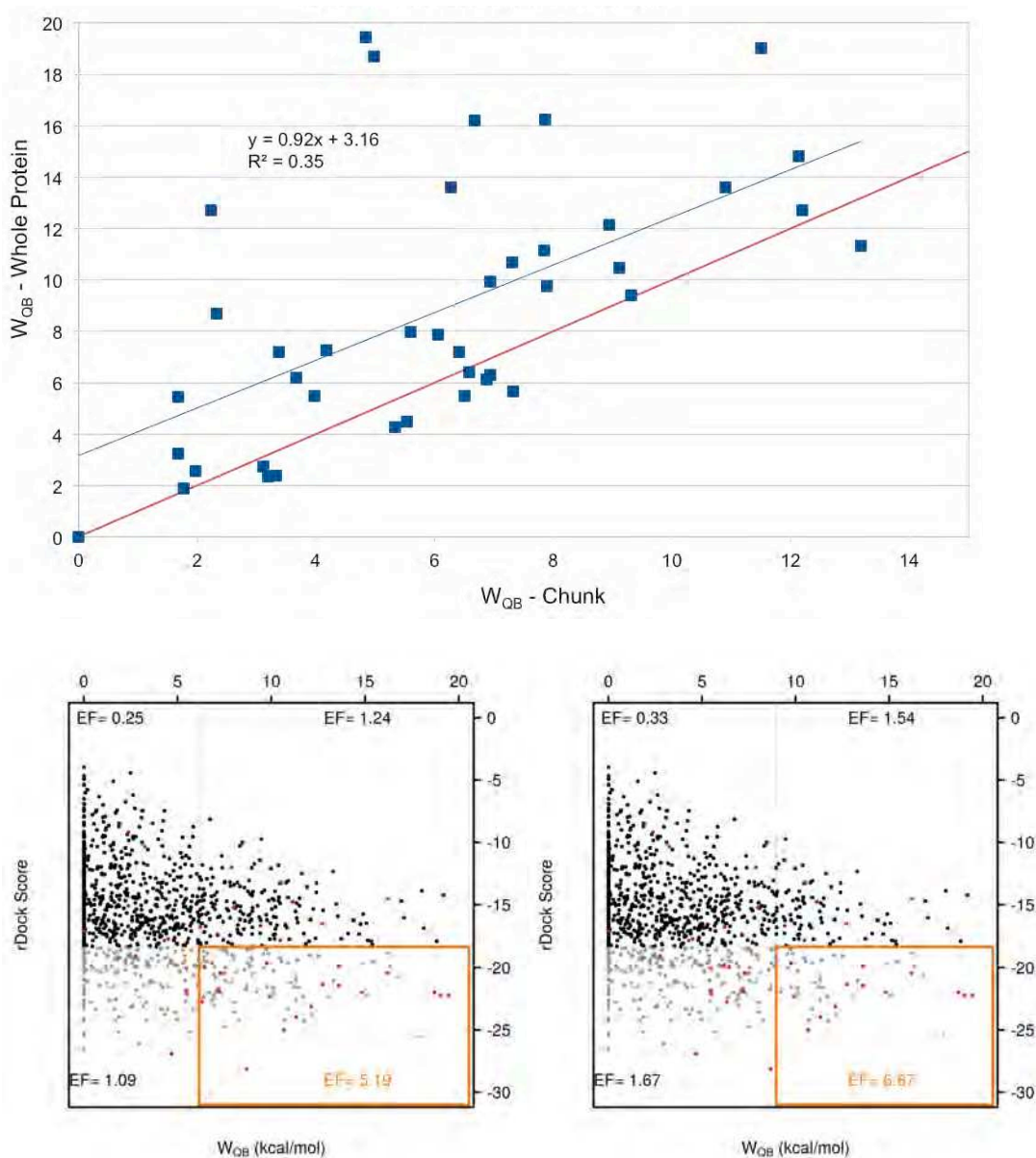
Supplementary Figure 14

Predicted binding modes for compounds **4**, **5** and **6** (from left to right).



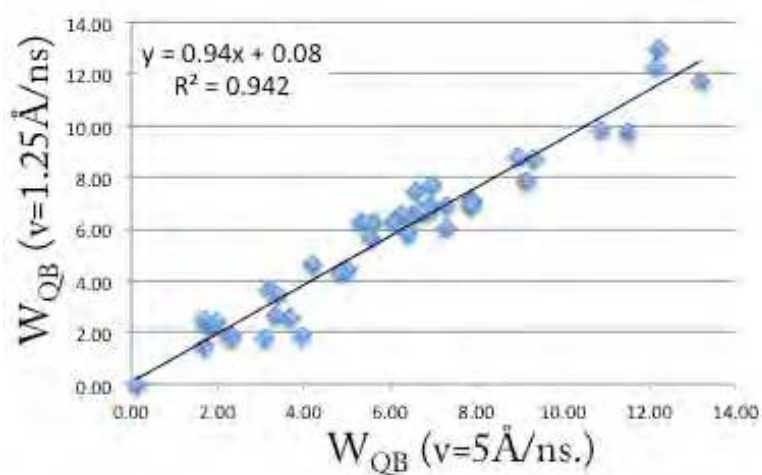
Supplementary Figure 15

Dependence of the results on the size of the receptor. W_{QB} values of CDK2 ligands were calculated using the whole protein as receptor and plotted against the results obtained with a truncated system (top). W_{QB} values obtained with the truncated system represent a lower bound to those obtained with the full system. This indicates that when the whole system is included, W_{QB} may not reflect the contribution of the interaction under investigation. Potentially, this may give rise to false positives. Noteworthy, the virtual screening results are comparable to those obtained with the truncated system (bottom; compare with Supplementary Figure 8).



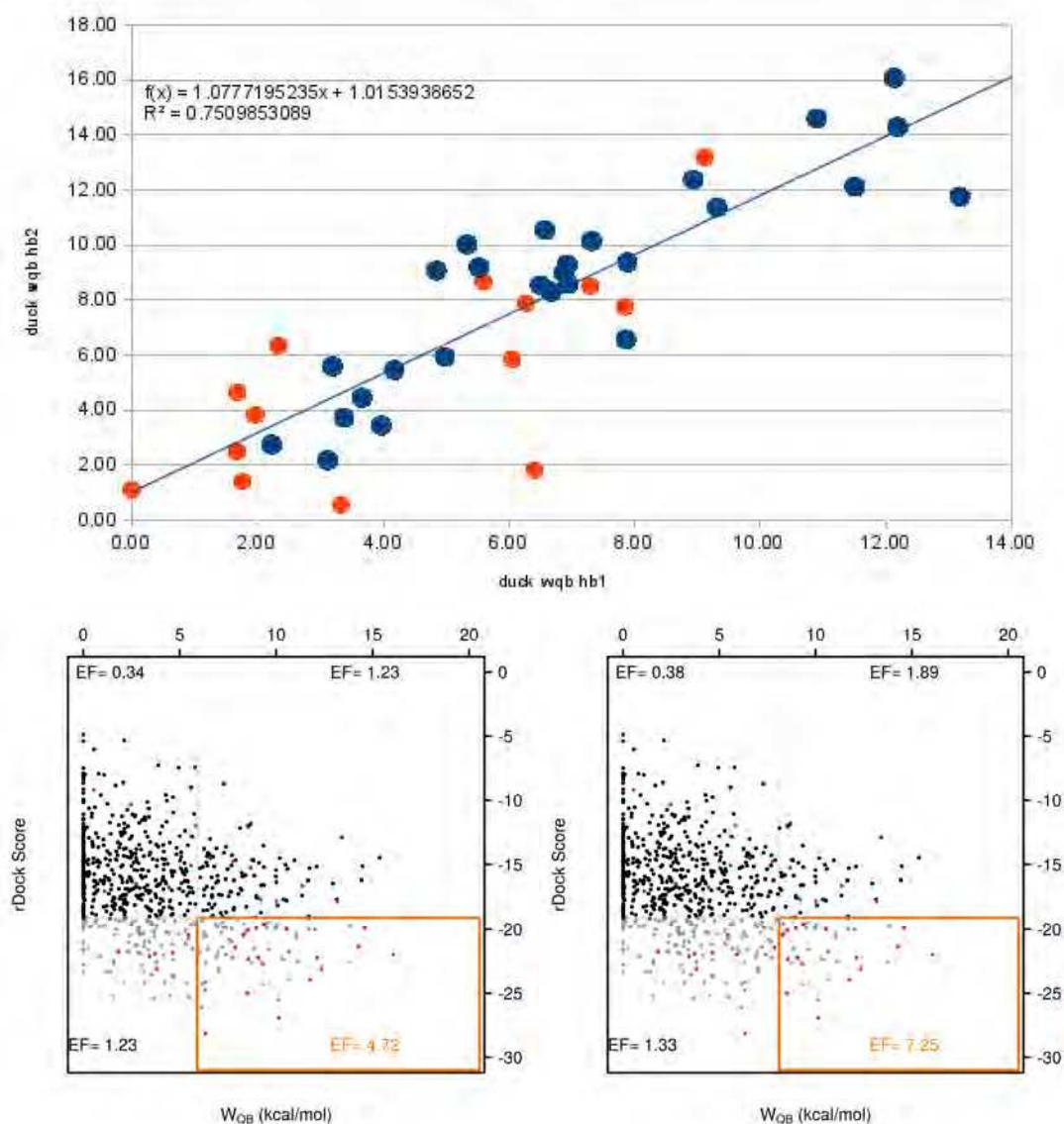
Supplementary Figure 16

Dependence of the results on the steering velocity. Two different velocities are compared: $5 \text{ \AA}\cdot\text{ns}^{-1}$ (used through this work) and $1.25 \text{ \AA}\cdot\text{ns}^{-1}$. Slower velocities mean more sampling and, potentially, lower W_{QB} values. The high correlation ($r^2=0.94$) indicates that the standard conditions ($v=5 \text{ \AA}\cdot\text{ns}^{-1}$) produce converged results.



Supplementary Figure 17

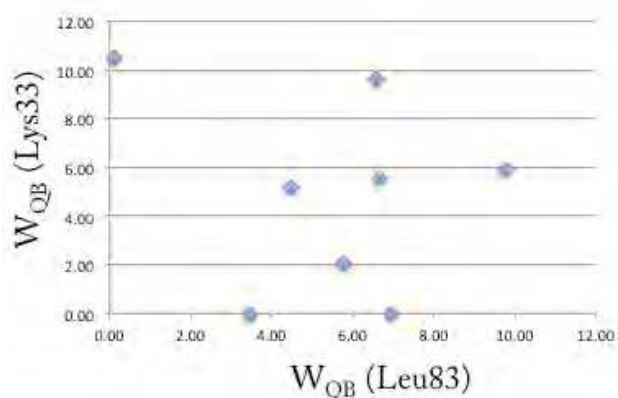
Dependence of the results on the choice of reaction coordinate. W_{QB} values obtained using two different atoms of reference in the hinge region of CDK2 are highly correlated (top) and afford similar enrichment factors in retrospective virtual screening (bottom; compare with Supplementary Figure 8). The atoms used as reference (Leu83:N in the x-axis and Leu83:O in the y-axis) are part of the hinge and located in close proximity (3Å). Most ligands form a hydrogen bond with both atoms at the same time. Points in red represent ligands that only form a hydrogen bond with Leu83:N.



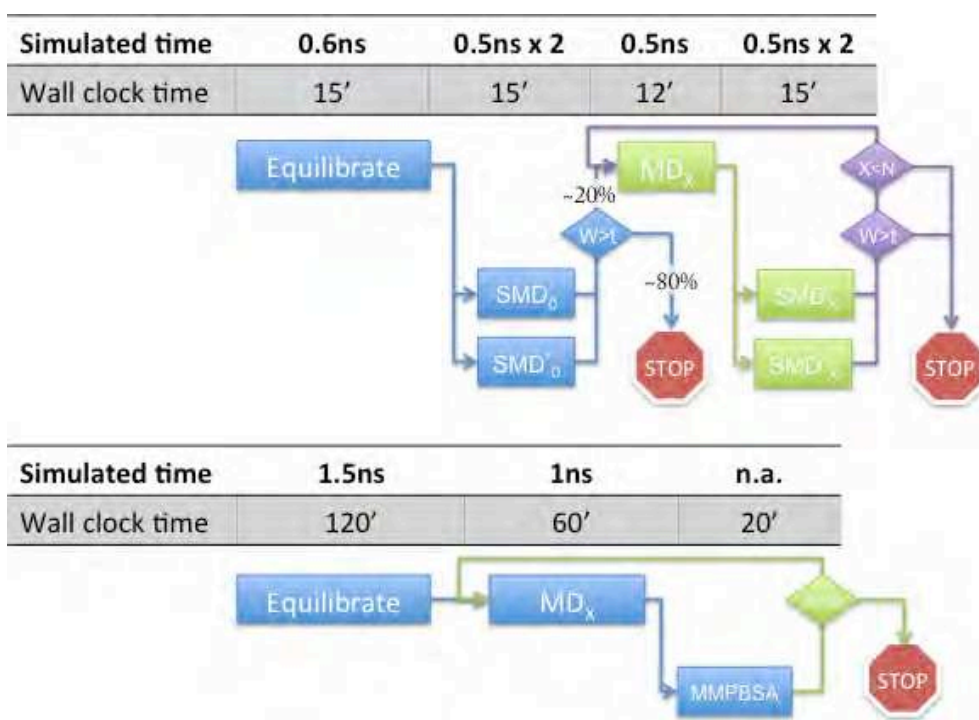
Supplementary Figure 18

W_{QB} of CDK2 ligands pulling the amine of Lys33 and comparison with W_{QB} values obtained for the hinge region (in kcal/mol). Only those ligands capable of forming a hydrogen bond with NZ of Lys33 have been considered. It should be noted that this part of the active site presents large conformational diversity between structures. In consequence, the DUCK results may be less reliable than for the hinge region.

PDB Code	W_{QB} (O Leu83)	W_{QB} (N ζ Lys33)
1OIQ	4.48	5.18
3BHT	6.59	9.65
3BHV	6.94	0.00
3EJ1	5.77	2.06
3FZ1	0.12	10.50
3QTQ	6.66	5.56
3QTW	9.76	5.91
3TIY	3.47	0.00

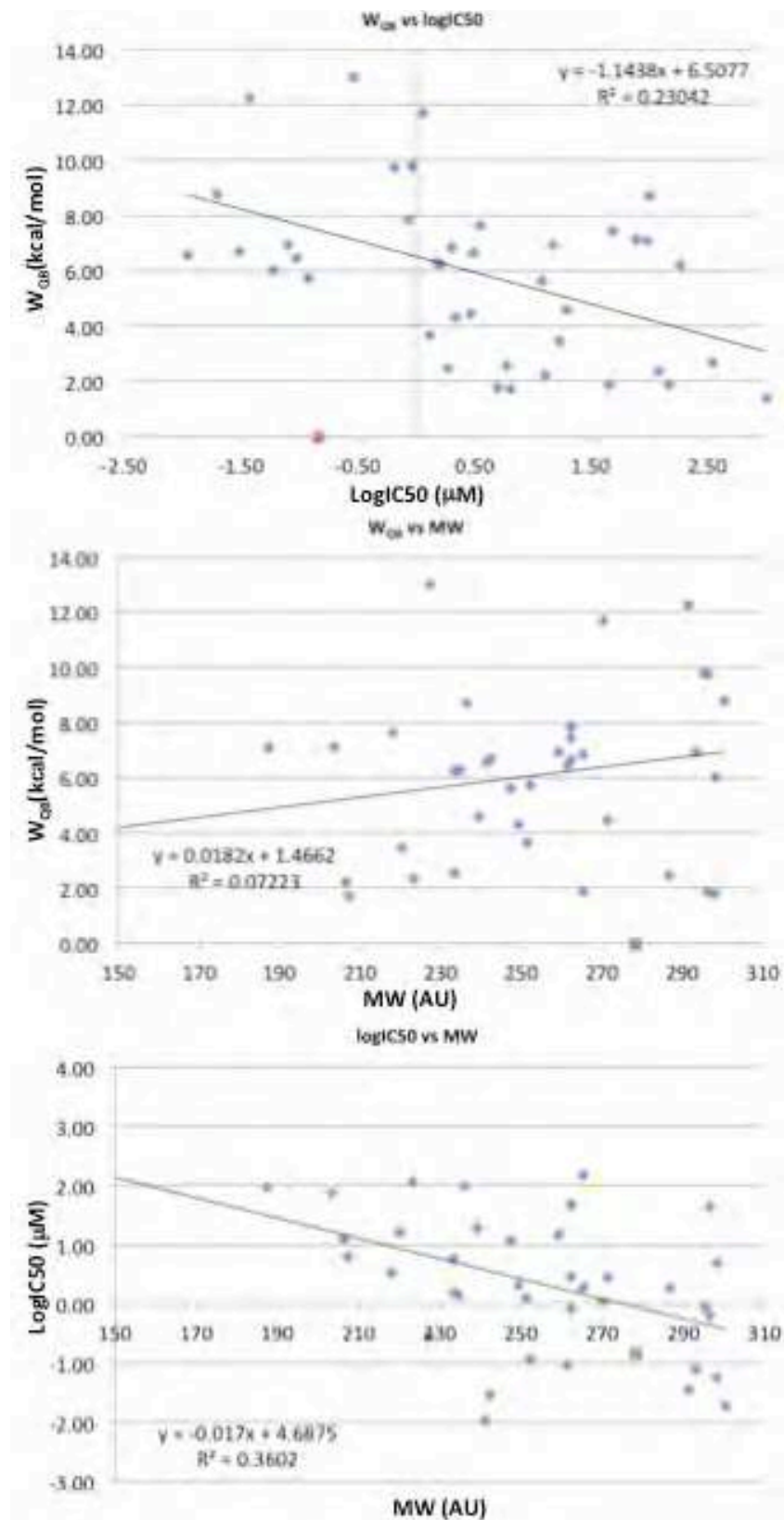

Supplementary Figure 19

Proposed protocol for DUCK-based virtual screening and comparison with MMPBSA. The smaller size of the system speeds up calculations by a factor of 5 (Supplementary Table 2), also permitting shorter equilibration times. Each ligand undergoes equilibration and at least two SMDs (45 GPU minutes). Molecules with W_{QB} above a given threshold (e.g. $t=6$ kcal/mol) would then proceed to N cycles of unbiased MD + SMD simulations (42 GPU minutes per cycle). A similar protocol for MMPBSA would require at least 2 GPU hours of equilibration followed by N cycles of 1ns MD simulation and MMPBSA calculation of representative snapshots (1 GPU hour + 20 CPU minutes per cycle).



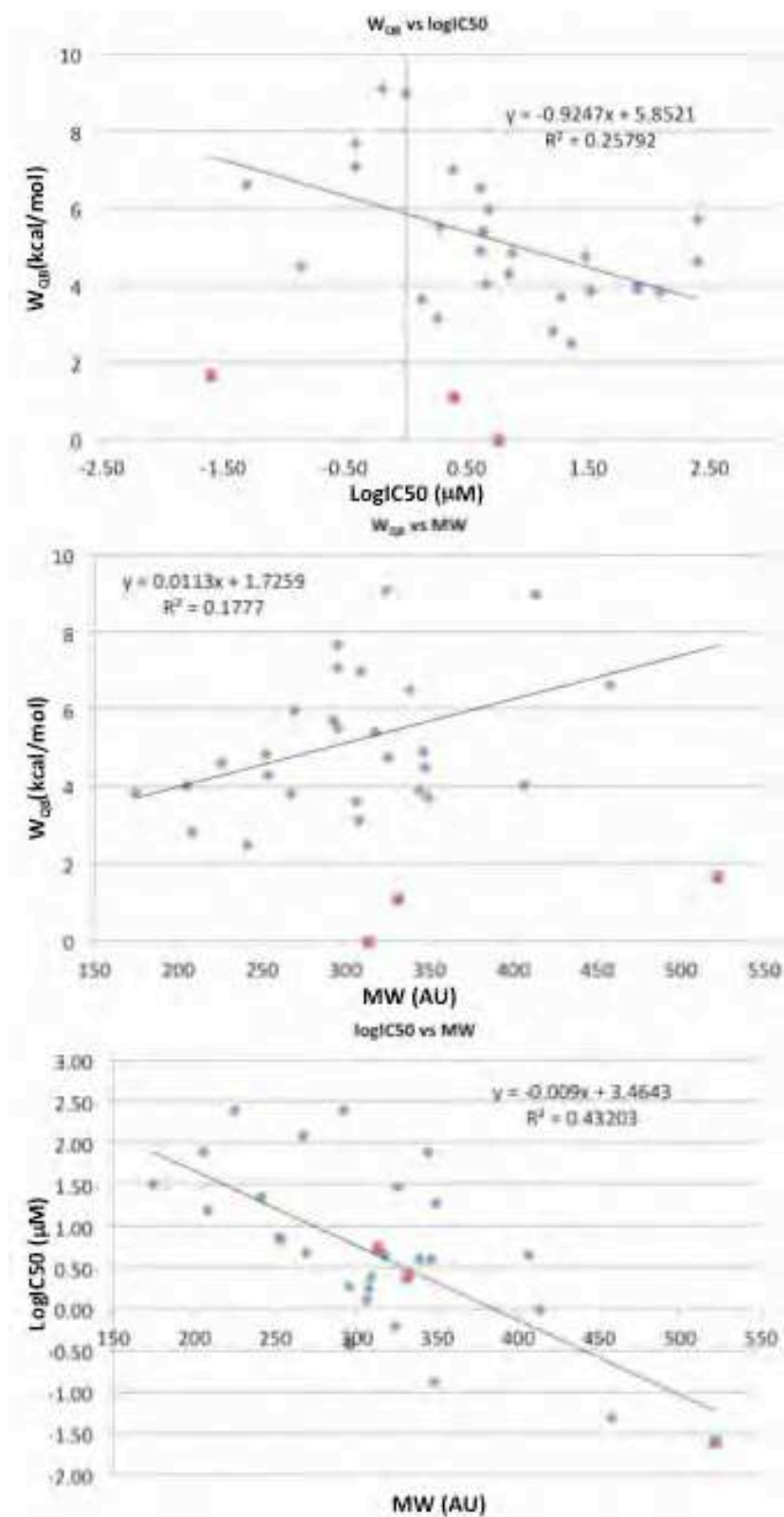
Supplementary Figure 20

CDK2 test set: correlation between LogIC_{50} and W_{QB} is not caused by Molecular Weight. The correlation between W_{QB} and MW ($r^2=0.07$) is lower than the correlation between W_{QB} and LogIC_{50} ($r^2=0.23$) or between LogIC_{50} and MW ($r^2=0.36$). Red points (discussed in Supplementary Figure 2) are excluded from all correlations.



Supplementary Figure 21

BRD4 test set: correlation between LogIC_{50} and W_{QB} is not caused by Molecular Weight. The correlation between W_{QB} and MW ($r^2=0.17$) is lower than the correlation between W_{QB} and logIC_{50} ($r^2=0.26$) or between LogIC_{50} and MW ($r^2=0.43$). Red points (discussed in Supplementary Figure 5) are excluded from all correlations.

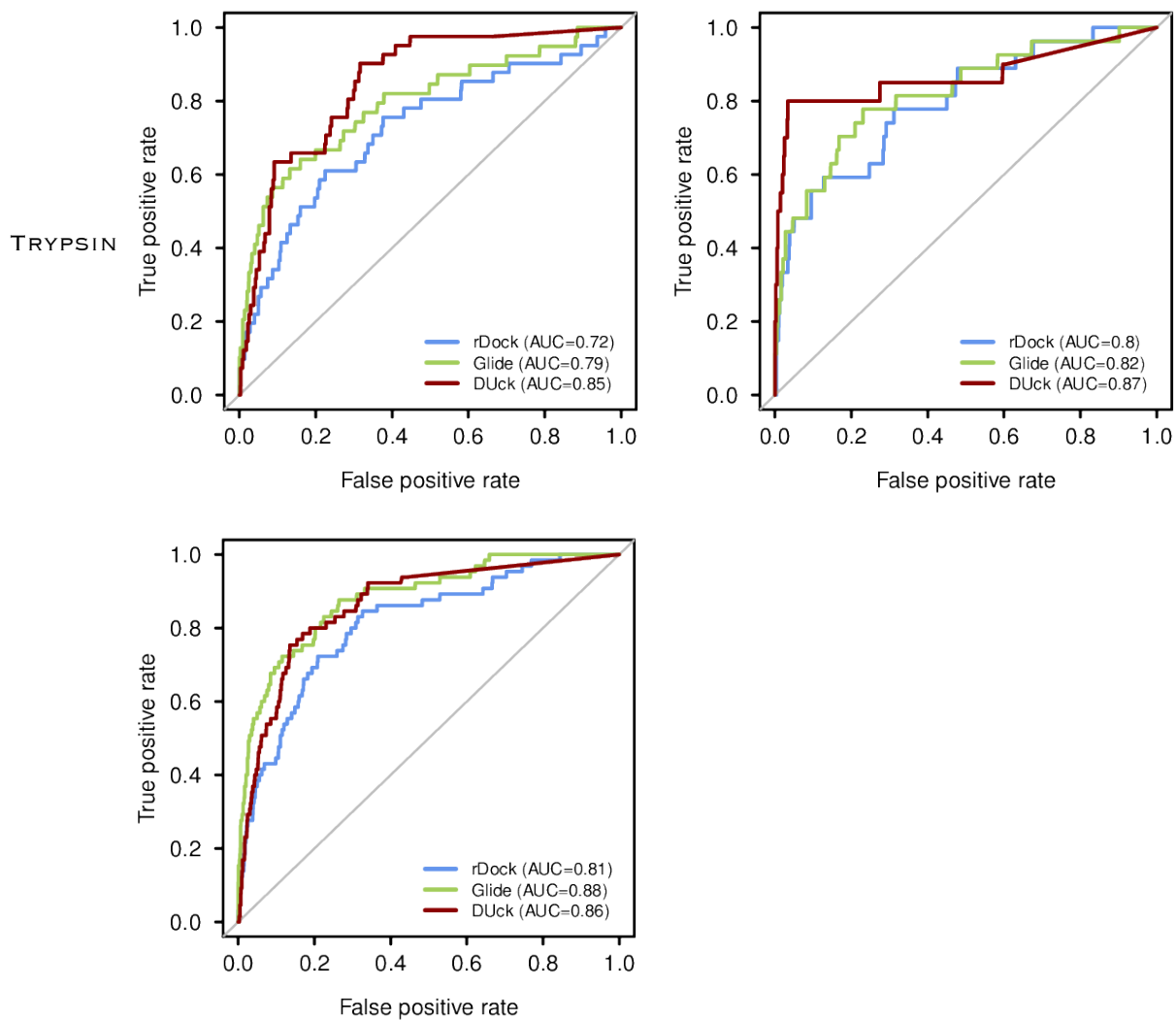


CDK2

AA2AR

Supplementary Figure 22

ROC curves comparison of DUck (in standalone mode) with unbiased docking with Glide and rDock for the three test systems: CDK2, AA2AR and Trypsin.

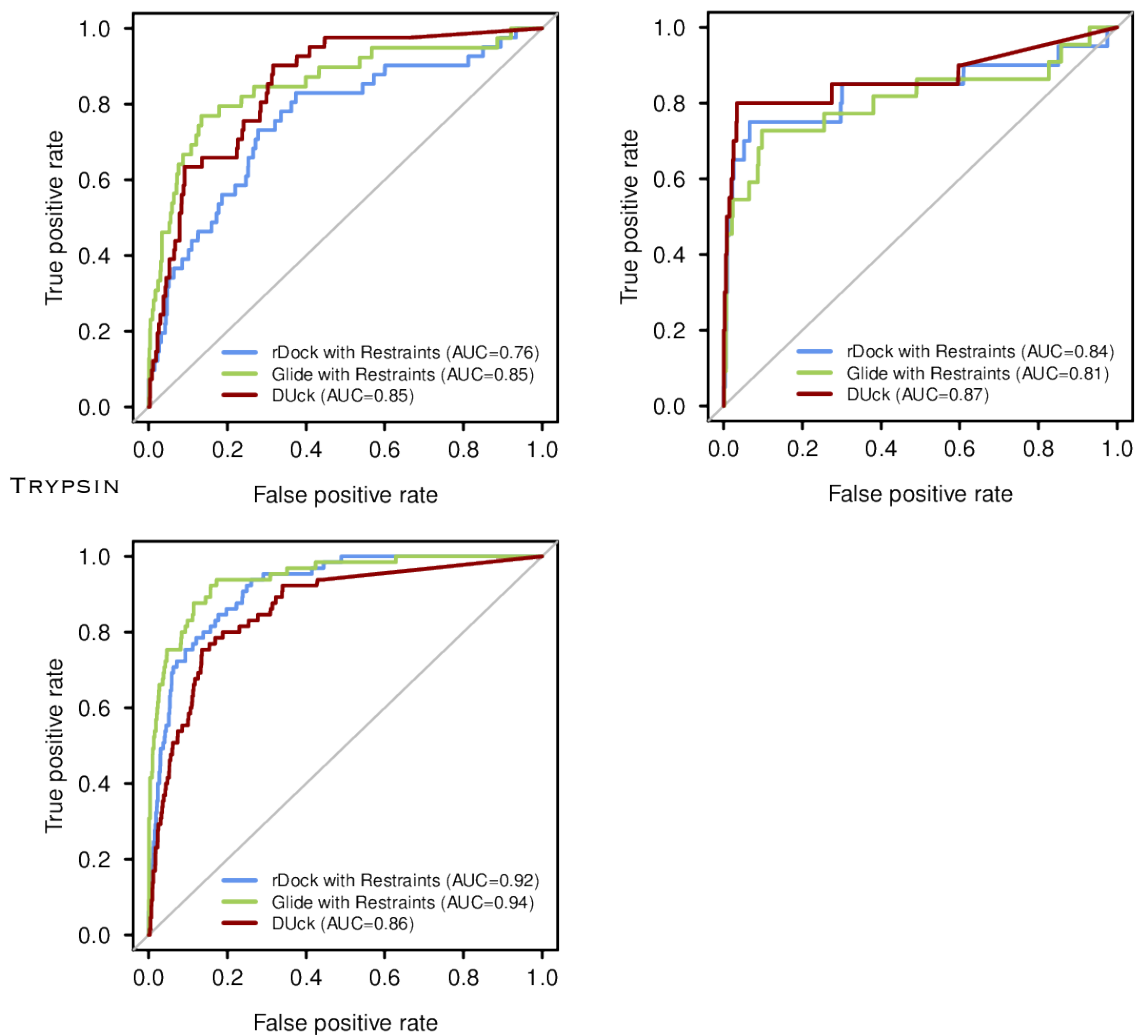


AA2AR

CDK2

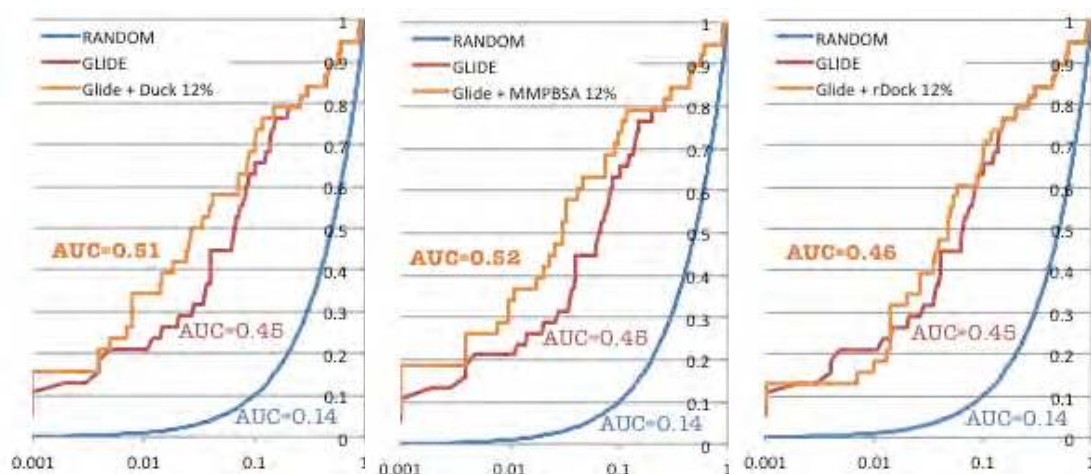
Supplementary Figure 23

ROC curves comparison of DUck (in standalone mode) with pharmacophore-guided docking with Glide and rDock for the three test systems: CDK2, AA2AR and Trypsin.

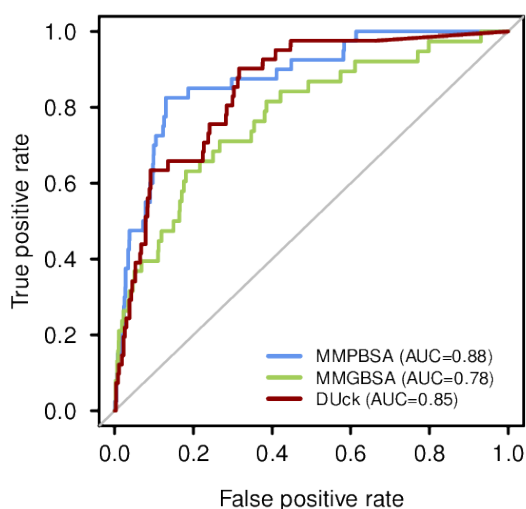


Supplementary Figure 24

DUck postfiltering improves early enrichment. Semilogarithmic ROC curves for the retrospective virtual screening of CDK2, obtained with the best-performing program for this test set (Glide), alone or in combination with three different postfiltering methods: DUck (left); MMPBSA (middle) and rDock (right). Ligands were initially ranked according to Glide's score. Then, moved to the back of the list if they were not in the top 12% of the rescoring method. This shows that the Glide-DUck combination is superior to Glide alone. For this test set the effect is most prominent in the top 1% to 5% of the library. Glide-MMPBSA combination is provided for comparison and affords very similar results. The Glide-rDock combination does not improve early enrichment.

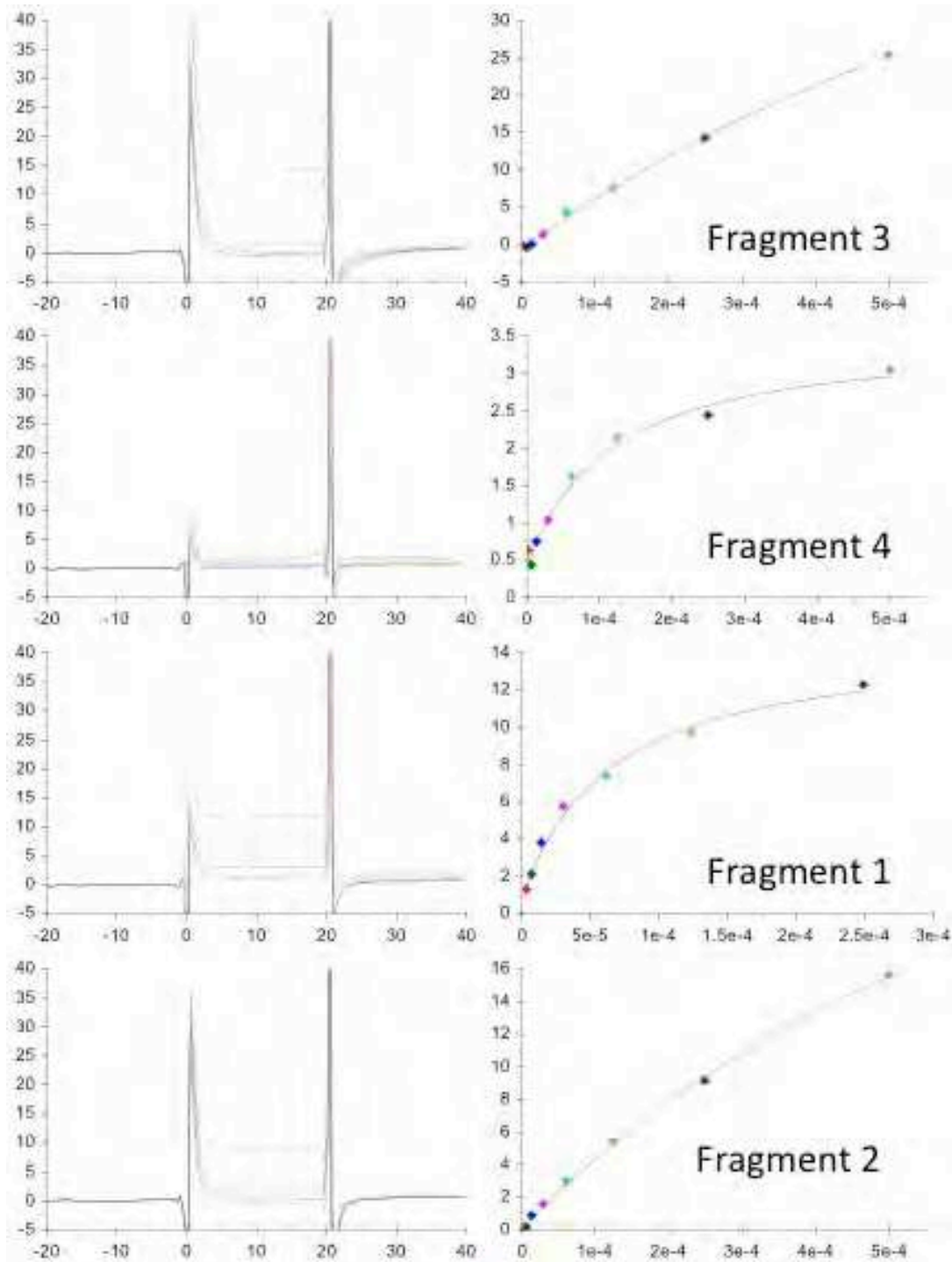
**Supplementary Figure 25**

ROC curves comparison of DUck (in standalone mode) with MMPBSA and MMGBSA for CDK2



Supplementary Figure 26

Examples of typical sensorgrams (left column) and steady state plots (right column) for the binding of the fragment hits to Hsp90. Fragments were tested in a 2-fold dilution series starting at 500uM or 250uM concentrations. Steady state values were calculated 4seconds before the injection stopped and plotted against the concentration. The K_D value was calculated by fitting the data to a steady state affinity model (Biacore T200 evaluation software GE Healthcare)



SUPPLEMENTARY TABLES

Supplementary Table 1

Chemical structures and summary of results for the 9 Hsp90 NMR Class 1 hits.

ID	Structure	MW	Docking		DUck		Xray	SPR Kd (mM)	PDB Sim ^b	ChEMBL Sim ^b
			Score	Rank ^a	W _{QB}	Rank ^a				
1		248,72	-24,97	79	9,1	10	Yes	77	2XDX (0.37)	CHEMBL 1340447 (0.44)
2		221,29	-25,03	73	8,2	11	Yes	320	2WI6 (0.29)	CHEMBL 1536318 (0.54)
3		230,22	-26,62	19	11,3	1	Yes	700	4EFU (0.32)	CHEMBL 1458840 (0.51)
4		240,33	-26,45	22	7,4	16	-	730	3WHA (0.29)	CHEMBL 1542436 (0.37)
5		165,20	-23,77	128	8,1	12	-	-	4EFT (0.27)	CHEMBL 1313412 (0.28)
6		206,27	-23,26	138	9,5	5	-	-	3HHU (0.42)	CHEMBL 2103879 (0.42)
7		236,30	-25,45	51	7,8	15	-	-	3B24 (0.31)	CHEMBL 1375884 (0.36)
8		224,66	-25,35	58	7,0	22	-	-	2XDX (0.35)	CHEMBL 1383799 (0.37)
22		237,29	-28,27	2	5,6	33	-	-	300I (0.27)	CHEMBL 1834092 (0.33)

^a Position within the list of 149 molecules that were evaluated with DUck. ^bHsp90 structure in the PDB or compound with Hsp90 activity in ChEMBL (as of 23/03/2016) with the closest similarity to the fragment hit. Similarity (values in parentheses) was calculated with Open Babel using the FP2 fingerprint.¹⁴

Supplementary Table 2

Number of atoms of the investigated systems. On average, using a protein chunk with explicit solvation produces a system 20% in size relative to the whole protein. As computational times scale linearly with the number of particles, this represents a 5-fold gain in efficiency.

Number of Atoms

System	Full System		Protein Chunk for DUck	
	Protein	Periodic Box ^a	Protein ^b	Periodic Box ^{a,b}
Hsp90	3291	30387	527 (16,0%)	9415 (31,0%)
Cdk2	4578	46803	345 (7,5%)	9110 (19,5%)
AA2AR	4603	73039	525 (11,4%)	8815 (12,1%)
Trypsin	3231	26721	335 (10,4%)	9696 (36,3%)
Average	3926	44238	433 (11,0%)	9259 (20,9%)

^a Protein solvated with TIP3 water molecules using Amber's tleap program. In all cases, the periodic system is a truncated octahedral box, the distance parameter is 12.0 and the closeness parameter is 0.65. ^b Values in parentheses are percentage of atoms relative to the full system.

Supplementary Table 3

Detail of the receptor definition used in DUck simulations. Water and residue numbers were taken from the corresponding PDB file.

System	Reference Atom	PDB Code (Chain)	Protein residues included as receptor	Water Molecules
CDK2	LEU 83 NH	1CKP (A)	ILE10 VAL18 LYS20 ALA21 VAL29 VAL30 ALA31 LEU32 VAL64 PHE80 GLU81 PHE82 LEU83 HIS84 GLN85 ASP86 LEU133 LEU134 ILE135 ASN136 ALA144	-
AA2AR	ASN 253 ND2	3EML (A)	LEU167 PHE168 GLU169 VAL172 PRO173 MET174 MET177 VAL178 ASN181 PHE182 TRP246 LEU247 PRO248 LEU249 HIS250 ILE251 ILE252 ASN253 CYS254 PHE255 THR256 PHE257 HIS264 ALA265 PRO266 LEU267 MET270 TYR271 LEU272 ALA273 ILE274	-
Trypsin	ASP189 OD1	2AYW (A)	HIS57 LEU99 ASP102 ASP189 SER190 CYS191 GLN192 GLY193 ASP194 SER195 VAL213 SER214 TRP215 GLY216 SER217 GLY219 CYS220 ALA221A GLN221 LYS224 PRO225 GLY226 VAL227 TYR228 THR229	1017 1096 1098 1101
Hsp90	ASP93 OD2	2YED (A)	GLU47 LEU48 ILE49 SER50 ASN51 SER52 SER53 ASP54 ALA55 LEU56 ASP57 LYS58 ILE78 ILE91 VAL92 ASP93 THR94 GLY95 ILE96 GLY97 MET98 GLY137 PHE138 VAL150 ILE151 THR152 GLY183 THR184 LYS185 VAL186	2043 2045 2049 2105 2107
BRD4	ASN140 ND2	3U5L (A)	TRP81 PRO82 PHE83 GLN84 GLN85 PRO86 VAL87 ASP88 ALA89 LYS91 LEU92 ASN93 LEU94 TYR97 ILE101 PRO104 MET105 THR131 ASN135 CYS136 TYR137 TYR139 ASN140 ASP144 ASP145 ILE146 MET149	-

Supplementary Table 4

Data collection and refinement statistics for Hsp90 in complex with Compounds **1**, **2** and **3**. R_{free} is the R factor calculated using 5% of the reflection data chosen randomly and omitted from the refinement process, whereas R_{cryst} is calculated with the remaining data used in the refinement. Rms bond lengths and angles are the deviations from ideal values; the rms deviation in B factors is calculated between covalently bonded atoms.

Compound	1	2	3
<i>Data collection statistics</i>			
Resolution (Å)	2.20	2.00	2.10
Space group	I222	I222	I222
Cell dimensions (Å)			
a =	66.87	64.96	68.98
b =	90.29	88.41	88.18
c =	98.33	99.06	96.90
No. molecules/asymmetric unit	1	1	1
Solvent content (%)	57.25	54.73	57.41
Measured reflections	66152	66886	62479
Unique reflections	15401	19011	17526
Completeness: Overall / in hrb ^a (%)	99.5 / 98.5	96.7 / 90.9 ^b	99.4 / 99.9
Mean I/σI: Overall / in hrb	11.2 / 2.8	11.1 / 1.3	8.33 / 0.95
R_{merge} : Overall / in hrb (%)	0.083 / 0.315	0.048 / 0.412	0.074 / 0.555
<i>Refinement statistics</i>			
R_{free} (%)	24.0	30.8 ^b	27.6
R_{cryst} (%)	19.1	22.1	22.4
Rms Deviations:			
Bonds (Å)	0.018	0.019	0.019
Angles (°)	1.920	1.958	2.046
B Factor (Å ²)	4.679	5.536	6.415
PDB Code	5FNC	5FND	5FNF

^ahrb: highest resolution bin. ^bDiffraction data for this structure was collected from a crystal that did not cryo-freeze correctly therefore the data in some of the resolution bins was of a lesser quality than the equivalent data collected from the other two crystals. This is likely to have impacted the refinement statistics for this structure.

Supplementary Table 5

List of ligands in the CDK2 test set. Ligands highlighted in red are not included in the correlation plotted in Supplementary Figure 2 and Supplementary Figure 25.

PDB	No.Atoms	MW	IC50 (uM)	Log IC50	W _{QB} (kcal/mol)
1E1V	21	247.303	12.00	1.08	5.53
1E1X	22	251.292	1.30	0.11	3.19
1JSV	23	265.293	2.00	0.30	6.93
1JVP	19	233.274	1.60	0.20	5.59
1OIQ	23	271.325	2.90	0.46	4.98
1PF8	21	242.26	0.03	-1.51	6.88
1PXJ	16	206.267	13.00	1.11	1.76
1PXK	19	249.293	2.20	0.34	4.84
1PXM	23	298.365	0.06	-1.22	7.32
1VYW	24	291.355	0.04	-1.43	12.13
1VYZ	19	227.268	0.29	-0.54	12.19
1W0X	25	298.351	5.00	0.70	3.97
1WCC	10	129.55	350.00	2.54	3.33
2BTR	19	261.344	0.10	-1.02	6.50
2BTS	24	300.417	0.02	-1.70	8.94
2C4G	22	270.294	1.15	0.06	13.18
2C5O	17	207.275	6.50	0.81	3.12
2CLX	21	218.221	3.50	0.54	6.94
2EXM	17	203.249	78.00	1.89	7.84
2R3H	19	239.282	20.00	1.30	4.18
2VTA	10	118.139	185.00	2.27	6.05
2VTH	18	223.249	120.00	2.08	1.97
2VTJ	22	286.739	1.90	0.28	1.68
2VTL	16	187.203	97.00	1.99	7.89
2VTM	11	144.137	1000.00	3.00	1.68
2VTN	22	262.246	0.85	-0.07	9.11
2VTR	16	234.67	1.50	0.18	5.34
3BHT	20	241.255	0.01	-1.96	6.28
3BHV	26	293.291	0.08	-1.10	7.30
3EJ1	20	252.281	0.12	-0.92	6.41
3FZ1	24	278.352	0.15	-0.84	0.12
3PXY	22	233.233	5.90	0.77	3.66
3QQK	21	259.328	15.00	1.18	7.87
3QTQ	21	262.332	3.10	0.49	6.67
3QTR	24	295.361	0.93	-0.03	10.90
3QTW	24	296.349	0.65	-0.19	11.51
3R8Z	21	262.332	49.00	1.69	6.57
3RZB	20	236.294	100.00	2.00	9.31
3TIY	20	220.185	17.00	1.23	3.38
3TIZ	23	265.314	150.00	2.18	2.23
4EZ3	25	296.305	45.00	1.65	2.33

Supplementary Table 6

List of ligands in the BRD4 test set. Ligands highlighted in red are not included in the correlation plotted in Supplementary Figure 5 and Supplementary Figure 26.

PDB	No.Atoms	MW	IC50 or Kd (nM)	Log IC50	W_{QB} (kcal/mol)
3MXF	31	458.00	49.00	-1.31	6.63
3U5J	22	308.77	2460.00	0.39	7.00
3U5L	23	323.78	640.00	-0.19	9.12
4A9L	22	325.38	30000.00	1.48	4.78
4C66	23	343.85	79400.00	1.90	3.92
4CFK	23	307.35	1830.00	0.26	3.13
4CFL	23	306.36	1330.00	0.12	3.63
4E96	24	347.39	136.00	-0.87	4.51
4HBV	13	241.09	23000.00	1.36	2.51
4HBW	18	269.32	4800.00	0.68	5.98
4HBX	20	295.36	1900.00	0.28	5.53
4HBY	22	317.36	4400.00	0.64	5.42
4HXR	21	338.41	4100.00	0.61	6.54
4HXS	23	346.42	4100.00	0.61	4.90
4J0R	22	295.34	386.00	-0.41	7.70
4J0S	22	295.34	382.00	-0.42	7.10
4LR6	13	174.20	33000.00	1.52	3.85
4LZS	15	208.26	16000.00	1.20	2.84
4MEN	20	267.33	125000.00	2.10	3.84
4MEO	22	292.34	250000.00	2.40	5.73
4MEQ	17	225.25	250000.00	2.40	4.62
4O72	30	413.49	1000.00	0.00	9.00
4O74	38	521.66	25.00	-1.60	1.67
4O77	25	331.35	2500.00	0.40	1.10
4O78	30	406.44	4600.00	0.66	4.07
4O7A	23	349.17	19000.00	1.28	3.72
4O7E	24	313.36	5700.00	0.76	0.00
4PCE	19	253.34	7000.00	0.85	4.30
4PCI	19	252.31	7500.00	0.88	4.84
4UYD	15	205.22	79400.00	1.90	4.03