Protein Structures in Virtual Screening: A Case Study with CDK2

Mark P. Thomas,* Campbell McInnes, and Peter M. Fischer†

Cyclacel Ltd., James Lindsay Place, Dundee, DD1 5JJ, U.K.

Received June 13, 2005

The influence of protein structure on the successful reproduction of known ligand poses by high-throughput docking programs is rarely discussed. Two commonly used programs, Glide and GOLD, were used to dock a set of CDK2 inhibitors of known bound pose into 20 different CDK2 structures. The numbers of docked poses that reproduced the known pose are reported. Depending on the program and protein structure, 0.3%−96.2% of the ligands docked with the correct pose. Although it is not possible to say that any one structure is "the best" for virtual screening, there are some structures that are clearly better than others. The main determinants of this are the volume of the binding site into which the ligands are docked and the exact orientation of the residues forming the binding site.

Introduction

Many reports have been published describing, and extolling the virtues of, one or other docking and/or scoring algorithm (see, for example, refs 1−5). Other papers have reported a side-by-side comparison of two or more virtual screening programs (see, for example, refs 6−11). Still more publications have described consensus docking, consensus scoring, or a mix-and-match approach that combines the docking algorithm from one program with the scoring algorithm from another (see, for example, refs 12−16). Many aspects of virtual screening have been reviewed elsewhere.[17−24] To a greater or lesser extent all of these publications have tried to define "the best" method of reproducing crystallographically observed ligand poses and/or calculating theoretical binding constants (or docking scores) that bear some resemblance to (or correlation with) experimentally determined values.

One variable that is rarely discussed in the literature on virtual screening is the influence of protein structure on the success of dockings. In most, if not all, enzymes conformational change—ranging from small side chain rotations to domain shifts—occurs upon ligand binding, with different ligands inducing different protein conformational changes. This has an impact on the choice of protein structure for use in high-throughput docking: if a protein conformation is optimally adapted for interaction with one specific ligand (as it might generally be expected to be in a crystal structure), does this limit the utility of that structure when screening a database of thousands of ligands of diverse structure? One way around the problem is to incorporate protein flexibility into the docking program. Full protein flexibility is not used in high-throughput docking because, given current computer hardware, it is incompatible with a docking program being high-throughput. If full protein flexibility was permitted, it is likely that there would be little, if any, difference between the protein structures. One approach to incorporating some level of protein flexibility into virtual screening is "soft docking" (see, for example, ref 25), where some steric clashes between protein and ligand are allowed. By loosening the criteria for steric fit, this implicitly models receptor flexibility. A disadvantage is that only small conformational changes are addressed. Another way around this problem is to dock a set of ligands into several different structures of the same protein and take the hits from each structure. (This is not consensus docking, where only those ligands that are hits against all or most of the proteins are selected.) A third way around the problem is to find a protein structure that is relatively promiscuous; i.e., it is able to accommodate the docking of a wide range of diverse ligands. One problem with a promiscuous protein structure is that, in a screen of a large and diverse database, more false positives might be found than if a less promiscuous protein structure had been used. (This is simply because the extra ligands that are docking into the more promiscuous structure are unlikely all to be true positives.) In hit and lead discovery, however, this is probably less of a problem than that of false negatives, where potentially novel hits are missed, and it is reasonable to suppose that, as well as having more false positives, a more promiscuous structure would have fewer false negatives than a less promiscuous structure. Although other papers mention this subject in passing, only a few deal explicitly with the subject of protein structures in virtual screening.

Birch et al.[26] studied the affect of protein conformation on docking success by taking the crystal structures of 31 neuraminidase/ligand complexes, extracting the ligands, and then docking each of the ligands into each of the protein structures. The correct ligand pose, to within 1.5 Å rmsd, was found between 32% and 81% of the time depending on the protein structure. The authors concluded that a single protein structure could be used for virtual screening but that care must be taken to identify the protein structure that performs best against a wide variety of ligands.

Erickson et al.[27] took three different proteins for which multiple crystal structures were available and selected one holo structure of each protein. The ligand was extracted from this structure and docked back into it and into an apo structure, and into an "average" protein structure that was selected by taking the structure that had the binding site coordinates closest to the average of all the structures of that protein. Unsurprisingly, the ligand most closely approached the crystallographically observed pose when docked into the holo structure and was furthest away from this pose in the apo structure. This reflects the degree of conformational change upon ligand binding and suggests that a holo structure should be used in virtual screening.

McGovern and Shoichet[28] describe how a set of 95 000 small molecules (the MDL Drug Data Report) was docked into holo,

* To whom correspondence should be addressed. Phone: +44-(0)1382-206062. Fax: +44-(0)1382-206067. E-mail: mthomas@cyclacel.com.

† Present address: Centre for Biomolecular Sciences, School of Pharmacy, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

10.1021/jm050554i CCC: $33.50
Published on Web 12/03/2005

apo, and modeled conformations of 10 different proteins. The 95 000 molecules included at least 35 ligands for each of the 10 systems. In seven systems the best enrichment was achieved with the holo structure, in two systems with the apo structure, and in one system with the modeled structure. The main points to arise from this work were three-fold. First, the performance of a molecular docking screen depends on the particular conformation of the receptor used in the calculation. The factors identified as being important in this respect were the binding site volume and the exact positioning of side chains: even a small change in the position of a side chain could have a substantial influence on the enrichment achieved. Second, the crystallographically determined holo conformation is the one most likely to yield meaningful enrichment of known ligands from a database containing mostly decoy molecules. Third, if the holo structure is overspecialized, i.e., optimized for one particular (class of) ligand, the apo or modeled structures may provide better discrimination between true ligands and decoys.

There were no protein kinases among the enzymes studied by Birch et al.,[26] McGovern and Shoichet,[28] or Erickson et al.[27] Given their ubiquity, roles in disease states, and potential as drug targets, the work we report here examines if the conclusions reached in these studies can be reached by studying kinases, specifically, cyclin-dependent kinase 2 (CDK2). Is it possible to identify one structure that is "the best" (in terms of the rate of correct pose prediction) for virtual screening? (In this context "the best" is not quite synonymous with "most promiscuous" (as defined above) because "the best" looks for correct poses rather than any bound pose.) Ligand binding and pose prediction are only half of a virtual screening program. The second half is scoring and ranking the poses. Binding of ligands in the correct pose is likely to be important for scoring and ranking because it is unlikely that correct ranking could consistently be achieved with incorrect binding modes, although, as demonstrated below, binding in the correct pose does not guarantee correct ranking. More than 60 CDK2 crystal structures are publicly available at present, including complexes of CDK2 with a set of 2-anilino-4-(hetero)aryl pyrimidine inhibitors solved in house.[29,30] We describe the use of two commonly used programs, Glide and GOLD, to dock a set of 340 CDK2 inhibitors of known binding pose and activity into 20 different CDK2 crystal structures, and for each program/protein structure combination, we report the number of docked poses that reproduce the crystallographically observed pose and go on to identify those aspects of the protein structure that influence the rate of correct pose reproduction.

CDK2 is an enzyme involved in the regulation of the cell cycle. The CDK2 protein, by itself, is inactive. A basal level of activity is conferred on it by the binding of a cyclin partner protein, which forms an active heterodimeric complex. (There are at least nine cyclins; CDK2 is known to bind cyclins A and E.) Phosphorylation of CDK2 at Thr160 is required to yield the fully active complex. Binding of cyclin A and phosphorylation of Thr160 changes the structure of CDK2. The tertiary structure of CDK2 is similar to other kinases, with an N-terminal domain of about 80 residues linked via a hinge region of about 10 residues to the C-terminal domain of about 130 residues. The ATP binding site is found in the cleft between the two domains. ATP binds in the cleft of the monomeric protein, but the triphosphate moiety is aligned in a catalytically inactive conformation. Upon binding cyclin A structural modifications to CDK2 produce an optimized ATP binding site: movement of helices bordering the cleft allows catalytically important residues into the correct position in the active site. Lys33 and Glu51 move to coordinate the α-phosphate of ATP, and Asp145, Phe146, and Gly147 induce in the triphosphate moiety of ATP the correct conformation for catalytic transfer. Phosphorylation of Thr160 induces further structural changes that form the binding site for the peptide substrate. It is the effect of these structural changes (as well as those induced by the presence of inhibitors in the ATP binding site) upon successful reproduction of the crystallographically observed binding mode that is being studied in this paper.

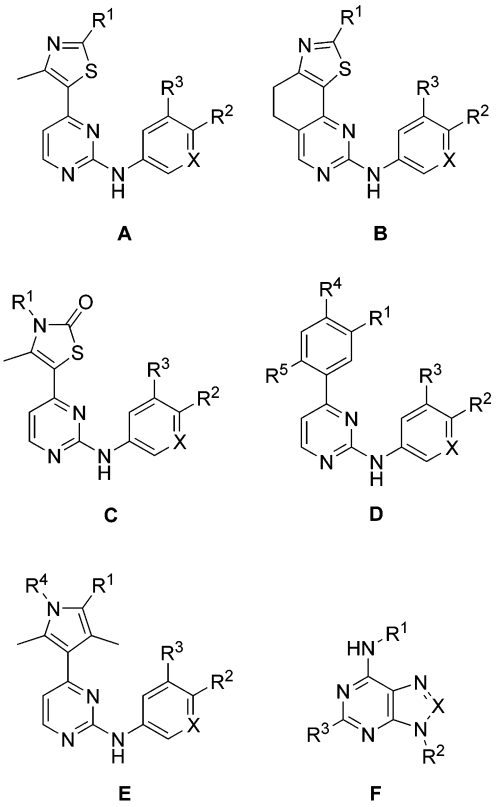

Figure 1. The core structures of group 1 (A), 2 (B−E), and 3 (F) compounds. X = CH, CR, or N.

Results

Docking Reproducibility. To compare results between different protein structures it was necessary to determine the extent to which the docking programs were able to reproduce the same results given the same initial conditions: if, given the same starting conditions, a program was unable to reproduce the same results (or produce similar results), it would be very difficult to compare the docking results for different proteins.

Glide is reported to use a Monte Carlo method in its docking algorithm [Glide documentation (Schrodinger, Inc.)[31]]. This random factor made it necessary to assess the reproducibility of Glide results. The set of 132 compounds in group 1 (Figure 1A) was docked into the 1AQ1, 1B38, 1DM2, 1E1V, 1FVV(a), 1FVV(c), 1GZ8, 1HCK, 1HCL, and 1JVP structures in both the standard and extra precision modes. This was done twice. Similarly, for the same protein structures, the 14 508 conformers of group 1 were rigidly docked twice. For a given operating mode and protein structure, it was found that the same ligands, docked under identical conditions, produced identical results (not shown). Discussion with the Glide scientists (Schrodinger, Inc.) revealed that Glide does not use a random number generator to generate trial moves in its pose refinement procedure and that, therefore, references to Monte Carlo are, regrettably, a little confusing. Glide dockings are, therefore,

Table 1. Number of Group 1 Compounds, out of 132, That Were Correctly Docked by GOLD in Five Docking Runs into 10 CDK2 Structures in the Library Screening (LS) Mode (on the left) and in the Standard Default (SD) Mode (on the right)

	number of successful dockings											
	LS mode						SD mode					
structure	run 1	run 2	run 3	run 4	run 5	average	run 1	run 2	run 3	run 4	run 5	average
1AQ1	104	100	102	102	97	101.0	81	79	84	84	79	81.4
1B38	35	34	32	33	37	34.2	60	64	64	63	64	63.0
1DM2	74	90	75	69	77	77.0	85	85	84	85	89	85.6
1E1V	86	83	78	82	82	82.2	124	123	125	125	128	125.0
1FVV(a)	119	120	124	116	120	119.8	129	126	128	126	125	126.8
1FVV(c)	112	112	107	112	110	110.6	127	126	124	128	126	126.2
1GZ8	57	60	61	60	58	59.2	119	120	122	122	121	120.8
1HCK	66	77	66	72	66	69.4	105	108	106	104	106	105.8
1HCL	41	34	33	41	37	37.2	126	123	127	128	121	125.0
1JVP	87	94	92	94	93	92.0	103	107	106	110	105	106.2

Table 2. Number of Group 1 Ligands That Were Correctly Docked by GOLD *n* Times in Five Docking Runs[a]

	number of ligands successfully docked *n* times											
	LS mode						SD mode					
structure	$n = 5$	$n = 4$	$n = 3$	$n = 2$	$n = 1$	$n = 0$	$n = 5$	$n = 4$	$n = 3$	$n = 2$	$n = 1$	$n = 0$
1AQ1	74	18	14	7	7	12	71	6	6	3	4	42
1B38	11	8	9	12	33	59	43	14	5	8	13	49
1DM2	46	20	11	14	14	27	64	17	10	3	4	34
1E1V	53	21	10	10	12	26	118	6	2	2	1	3
1FVV(a)	104	15	2	6	1	4	122	5	1	0	1	3
1FVV(c)	89	16	7	10	3	7	120	5	2	2	1	2
1GZ8	21	19	21	14	24	33	106	12	4	4	6	0
1HCK	27	24	18	18	26	19	88	14	6	4	7	13
1HCL	5	12	17	12	38	48	115	7	4	5	0	1
1JVP	58	27	11	6	17	13	90	13	5	4	6	14

[a] For example, in the library screening (LS) mode, over five runs, an average of 101 ligands out of 132 correctly docked into the 1AQ1 structure (Table 1). This table shows that 74 ligands were correctly docked in all five runs, 18 ligands were correctly docked in four out of the five runs, 14 in three runs, 7 in two runs, 7 in 1 run, and 12 ligands failed to dock correctly in any of the five runs. If GOLD were totally consistent, only the $n = 5$ and $n = 0$ columns would contain values greater than 0, as each ligand would be either correctly docked in every run or incorrectly docked in every run. For each protein structure, the total of $n = 5$ plus $n = 0$ is greater in the standard default (SD) mode than in the LS mode, showing that the SD mode produces more consistent results.

purely deterministic: given identical starting conditions Glide will produce identical results. Glide results are reproducible.

GOLD uses a genetic algorithm in its docking/scoring function. The randomness associated with genetic algorithms means that multiple docking runs under identical conditions are highly unlikely to produce identical results. The 132 ligands of group 1 were docked five times into the same 10 protein structures used to assess result reproducibility in Glide. This was done in both the library screening (LS) and standard default (SD) docking modes. The results are presented in Table 1, which shows that the numbers of ligands correctly docked are fairly consistent for any one protein structure and docking speed, with the spread in the number of correctly docked ligands ranging from 5 (1GZ8) to 16 (1DM2) in the LS mode and from 4 (1GZ8) to 8 (1HCL and 1JVP) in the SD mode. Table 1 does not, however, indicate whether it is the same ligands that are correctly docked in each run. This is shown in Table 2. Ideally, since every ligand is a known CDK2 inhibitor, every ligand would adopt the correct pose every time it is docked. This, however, is probably an unrealistic expectation. A more realistic expectation might be that each ligand would be either correctly docked in every run or incorrectly docked in every run. The results show that in neither the LS nor the SD modes with any of the 10 protein structures does GOLD achieve either of these states. It can be seen, however, that, in this respect, dockings in the SD mode are better than those in the LS mode. Ligands are more likely to dock either consistently correctly or consistently incorrectly in the SD mode than they are in the LS mode.

The set of 340 ligands was docked into 20 different CDK2 structures. Tables 3 and 4 show the docking results for Glide and GOLD, respectively, while Table 5 sums the results.

Glide Rigid Docking. The 38 684 ligand conformers were docked into the 20 protein structures to give a possible total of 773 680 docked poses. Of these, Glide found 134 227 (17.3%). Looking at the individual proteins, 1HCK returned the fewest poses (3384 out of 38 684, 8.7%) and 1FIN(c) the most (9176 out of 38 684, 23.7%). When the results are broken down by ligand group, it is unsurprising, given the similar structure of the ligands, that the group 1 (Figure 1A) and group 2 ligands (Figure 1B−E) should return similar results: totals of 14.1% and 15.1%, respectively, with, in both cases, 1HCK returning the fewest dockings (7.5% and 7.3% for group 1 and group 2 ligands, respectively) and 1FIN(c) returning the most dockings (group 1, 20.7%; group 2, 21.1%). Docking of the group 3 ligands (Figure 1F) returned a slightly higher overall docking rate (19.6%), with 1HCK again having the worst docking rate (10.8%) and 1G5S having the best (23.8%). The highest docking rates were achieved with the group 4 ligands (38.5%, Figure 2). At 17.4%, 1HCK was again the worst, and 1FIN(c) was the best with 49.8%.

Glide Flexible Docking: Standard Precision Mode. Across all 20 protein structures, 52.5% of the 340 ligands docked with the correct pose (as defined by formation of the requisite hinge-region hydrogen bonds and correct orientation of the ligand in the ATP binding site) when operating in the standard precision mode. Results for individual proteins ranged from 5.0% (1HCL) to 81.8% (1FIN(c)). The average rate of correct pose reproduction for the individual groups of compounds was 53.4% (group 1), 59.0% (group 2), 39.7% (group 3), and 34.6% (group 4). The 1HCL structure gave the lowest rate of correct pose reproduction in ligand groups 1, 2, and 4 with 3.8%, 5.0%, and 8.3%, respectively, and in group 3 the lowest rate of correct

Table 3. Number of Ligands Correctly Docked by Glide in the Standard Precision (SP) and Extra Precision (XP) Modes and the Number of Ligands Docked in the Rigid Mode

	group 1 (132 ligands, 14508 conformers)			group 2 (141 ligands, 15323 conformers)			group 3 (31 ligands, 5620 conformers)			group 4 (36 ligands, 3233 conformers)			all (340 ligands, 38684 conformers)		
structure	SP	XP	rigid	SP	XP	rigid	SP	XP	rigid	SP	XP	rigid	SP	XP	rigid
1AQ1	76	110	2402	79	117	2383	9	11	1125	12	13	1273	176	251	7183
1B38	9	2	1562	23	7	1614	14	22	1009	9	10	1064	55	41	5249
1DI8	103	106	2818	123	126	3000	17	14	1283	15	12	1516	258	258	8617
1DM2	80	33	1869	114	87	2454	5	6	1085	14	11	1244	213	137	6652
1E1V	42	60	1380	49	75	1646	0	0	1022	9	4	1105	100	139	5153
1E1X	53	61	1648	76	95	2148	4	3	1012	15	4	1207	148	163	6015
1FIN(a)	103	124	2781	103	109	2958	21	29	1328	18	16	1563	245	278	8630
1FIN(c)	121	117	3000	124	120	3240	18	27	1327	15	13	1609	278	277	9176
1FVT	87	74	2287	94	87	2647	7	14	1135	16	12	1308	204	187	7377
1FVV(a)	116	121	2213	114	114	2587	15	6	1100	16	15	1288	261	256	7188
1FVV(c)	123	126	2264	123	118	2597	15	23	1108	16	16	1309	277	283	7278
1G5S	81	109	2615	111	129	2695	18	25	1335	12	15	1471	222	278	8116
1GZ8	8	1	1764	42	9	1986	22	15	1075	9	4	1164	81	29	5989
1H00	94	105	2036	93	113	2359	26	29	1094	16	15	1242	229	262	6731
1H1P	38	19	1742	82	47	1998	4	0	1049	9	9	1224	133	75	6013
1HCK	20	27	1094	21	15	1124	11	22	605	8	7	561	60	71	3384
1HCL	5	0	1633	7	0	2125	2	0	1125	3	1	1184	17	1	6067
1JVP	87	108	2135	90	121	2328	13	5	1097	9	9	1228	199	243	6788
1OIT	110	121	2346	106	116	2607	19	25	1148	17	17	1297	252	279	7398
Rosc	55	77	1420	89	89	1714	6	13	1024	11	12	1065	161	191	5223
totals	1411	1501	41009	1663	1694	46210	246	289	22086	249	215	24922	3569	3699	134227

Table 4. Number of Ligands Correctly Docked by GOLD into 20 CDK2 Structures[a]

	group 1 (132 compounds)			group 2 (141 compounds)			group 3 (31 compounds)			group 4 (36 compounds)			all (340 compounds)		
structure	LS	SD	both	LS	SD	both	LS	SD	both	LS	SD	both	LS	SD	both
1AQ1	104	87	79	129	132	126	30	30	29	28	23	22	291	272	256
1B38	37	67	32	78	124	78	31	31	31	20	23	18	166	245	159
1DI8	100	71	66	119	123	108	25	31	25	30	27	25	274	252	224
1DM2	78	86	63	117	134	116	30	31	30	25	27	24	250	278	233
1E1V	89	125	89	101	122	100	21	31	21	21	24	17	232	302	227
1E1X	79	114	76	104	125	98	25	31	25	26	26	24	234	296	223
1FIN(a)	104	57	54	132	134	130	29	31	29	26	25	25	291	247	238
1FIN(c)	110	81	75	129	135	126	28	31	28	22	27	22	289	274	251
1FVT	102	105	95	123	139	122	29	31	29	28	28	26	282	303	272
1FVV(a)	123	129	122	135	138	134	29	31	29	25	29	24	312	327	309
1FVV(c)	109	127	108	129	137	126	25	31	25	27	26	24	290	321	283
1G5S	97	79	66	126	134	124	30	31	30	25	25	24	278	269	244
1GZ8	56	122	56	87	111	82	30	31	30	20	24	19	193	288	187
1H00	113	119	104	124	137	124	30	31	30	23	23	21	290	310	279
1H1P	48	70	40	87	121	82	19	31	19	22	19	14	176	241	155
1HCK	76	106	73	98	129	97	28	31	28	24	29	23	226	295	221
1HCL	40	126	40	59	120	57	5	31	5	12	28	12	116	305	114
1JVP	92	107	87	112	130	108	22	29	21	21	25	20	247	291	236
1OIT	88	111	84	119	138	118	26	31	26	27	30	27	260	310	255
Rosc	69	108	62	116	129	116	30	31	30	25	23	20	240	291	228
totals	1714	1997	1471	2224	2592	2172	522	617	520	477	511	431	4937	5717	4594

[a] The "LS" columns show the number of ligands correctly docked by GOLD operating in the library screening mode. The "SD" columns show the number of ligands correctly docked by GOLD operating in the standard default mode. The "both" columns show the number of ligands correctly docked by GOLD operating in both the LS and SD mode; e.g., of the 132 ligands of group 1 docked into 1AQ1, 79 were common to the 104 correctly docked in the LS mode and 87 docked in the SD mode.

pose reproduction was achieved by the 1E1V structure (0%). The highest rates of correct pose reproduction were with the following structures: 1FVV(c) (group 1, 93.2%), 1FIN(c) (group 2, 87.9%), 1H00 (group 3, 83.9%), and 1FIN(a) (group 4, 50.0%). Two further docking runs, using only the 132 compounds of group 1, were performed with the 1HCL and 1FIN(c) structures. The first was done after minimizing the position of the protein hydrogens and the second after minimizing the position of the protein side chains. The number of ligands docked in the correct pose in the unminimized, hydrogen-minimized, and side-chain-minimized 1HCL structures was 5 (3.8%), 5 (3.8%), and 0, respectively. The equivalent numbers for the 1FIN(c) structure were 121 (91.7%), 118 (89.4%), and 120 (90.9%).

Glide Flexible Docking: Extra Precision Mode. The rate of correct pose reproduction when operating in the extra precision mode was 54.4% for all 340 ligands docked into all 20 protein structures with, for the individual ligand groups, the following rates of correct pose reproduction being achieved: group 1, 56.9%; group 2, 60.1%; group 3, 46.6%; and, group 4, 29.9%. With no correct dockings, the 1HCL structure has the worst hit rate with the group 1 and group 2 ligands, as it does (along with the 1E1V and 1H1P structures) with the group 3 ligands, and with only one correct docking (2.8%) the 1HCL structure is worst with the group 4 ligands. This makes 1HCL the worst structure across all four groups (0.3% reproduction of correct poses). The highest rate of correct pose reproduction with the group 1 compounds was achieved with the 1FVV(c)

Table 5. Flexible Docking Totals[a]

	group 1			group 2			group 3			group 4			all		
structure	Glide	GOLD	total	Glide	GOLD	total	Glide	GOLD	total	Glide	GOLD	total	Glide	GOLD	total
1AQ1	186	191	377	196	261	457	20	60	80	25	51	76	427	563	990
1B38	11	104	115	30	202	232	36	62	98	19	43	62	96	411	507
1DI8	209	171	380	249	242	491	31	56	87	27	57	84	516	526	1042
1DM2	113	164	277	201	251	452	11	61	72	25	52	77	350	528	878
1E1V	102	214	316	124	223	347	0	52	52	13	45	58	239	534	773
1E1X	114	193	307	171	229	400	7	56	63	19	52	71	311	530	841
1FIN(a)	227	161	388	212	266	478	50	60	110	34	51	85	523	538	1061
1FIN(c)	238	191	429	244	264	508	45	59	104	28	49	77	555	563	1118
1FVT	161	207	368	181	262	443	21	60	81	28	56	84	391	585	976
1FVV(a)	237	252	489	228	273	501	21	60	81	31	54	85	517	639	1156
1FVV(c)	249	236	485	241	266	507	38	56	94	32	53	85	560	611	1171
1G5S	190	176	366	240	260	500	43	61	104	27	50	77	500	547	1047
1GZ8	9	178	187	51	198	249	37	61	98	13	44	57	110	481	591
1H00	199	232	431	206	261	467	55	61	116	31	46	77	491	600	1091
1H1P	57	118	175	129	208	337	4	50	54	18	41	59	208	417	625
1HCK	47	182	229	36	227	263	33	59	92	15	53	68	131	521	652
1HCL	5	166	171	7	179	186	2	36	38	4	40	44	18	421	439
1JVP	195	199	394	211	242	453	18	51	69	18	46	64	442	538	980
1OIT	231	199	430	222	257	479	44	57	101	34	57	91	531	570	1101
Rosc	132	177	309	178	245	423	19	61	80	23	48	71	352	531	883
totals	2912	3711	6623	3357	4816	8173	535	1139	1674	464	988	1452	7268	10654	17922

[a] For each protein structure and ligand group, this table shows the number of ligands correctly docked. Each ligand group is broken down by program: the "Glide" columns are the totals of docking in the SP and XP modes, and the "GOLD" columns are the totals of docking in the LS and SD modes.

structure (95.5%), and the 1G5S structure had the greatest number of correct pose reproductions for the group 2 ligands (91.5%). The 1FIN(a) and 1H00 structures both have a 93.5% rate of correct pose reproduction with the group 3 ligands, and the 1OIT structure is best with the group 4 ligands (47.2%). Overall, the 1FVV(c) structure has the best rate of correct pose reproduction (83.2%). Two further docking runs, using only the 132 compounds of group 1, were performed with the 1HCL and 1FIN(c) structures. The first was done after minimizing the position of the protein hydrogens and the second after minimizing the position of the protein side chains. The number of ligands docked in the correct pose in the unminimized, hydrogen-minimized, and side-chain-minimized 1HCL structures was 0 throughout. The equivalent numbers for the 1FIN(c) structure were 117 (88.6%), 120 (90.9%), and 121 (91.7%).

GOLD Docking: Library Screening Mode. Across all 20 protein structures and all the ligand groups, 72.6% of the ligands docked (in the LS mode) with the correct pose. The best and worst protein structures were 1FVV(a) (91.8%) and 1HCL (34.1%), respectively. The best structure with the group 1 ligands was 1FVV(a) (93.2%) and the worst was 1B38 (28.0%), with an average of 64.9%. 1FVV(a) was again the best structure with the group 2 compounds (95.7%), and 1HCL was the worst (41.8%) with an average of 78.9%. A 100% success rate was achieved when docking the group 3 compounds into 1B38 but only a 16.1% success rate with 1HCL. The average for the group 3 compounds was 84.2%. The group 4 compounds were docked with the highest success rate (83.3%) into the 1DI8 structure. The 1HCL structure had the lowest rate of correct pose reproduction (33.3%) and the average rate of correct pose reproduction across all 20 structures was 66.3%.

GOLD Docking: Standard Default Mode. In the SD docking mode, the 340 compounds were docked into the 20 protein structures with an average 84.1% rate of correct pose reproduction with values for individual proteins ranging from 70.9% (1H1P) to 96.2% (1FVV(a)). 1FVV(a) had the highest hit rate with the group 1 compounds (97.7%) and 1FIN(a) the lowest (43.2%) with an average of 75.6%. The highest rate of correct pose reproduction with the group 2 compounds was achieved with the 1FVT structure (98.6%) and the lowest with 1GZ8 (78.7%) and an average of 91.9%. With the group 3 compounds only the 1AQ1 (96.8%) and 1JVP (93.5%) structures achieved a success rate of less than 100%. The average across all 20 structures was 99.5%. The 1OIT structure had the highest rate of correct pose reproduction of the group 4 compounds (83.3%) and the 1H1P structure the lowest (52.8%) with an average of 71.0%.

GOLD: Both Modes. High-throughput docking programs are commonly used in a fast mode to screen a large database of compounds, the hits from which are then rescreened in a slow mode. Compounds that fail to dock in the fast mode will not get as far as screening in the slow mode and compounds that dock well in the fast mode but poorly in the slow mode are unlikely to be advanced to biochemical screening. Compounds that go forward for biochemical screening are likely to have docked well in both the fast and slow docking modes. In Table 4, the columns headed "both" list the number of compounds that were correctly docked in both the LS (fast) and SD (slow) docking modes. An average 67.6% of the 340 compounds docked well at both speeds with values ranging from 33.5% (1HCL) to 90.9% (1FVV(a)). The average, best, and worst rates of correct docking for the four groups of compounds are, respectively, as follows: group 1, 55.7%, 92.4% (1FVV(a)), and 24.2% (1B38); group 2, 77.0%, 95.0% (1FVV(a)), and 40.4% (1HCL); group 3, 83.9%, 100% (1B38), and 16.1% (1HCL); and group 4, 59.9%, 75.0% (1OIT), and 33.3% (1HCL).

The rightmost column of Table 5 lists, for each protein structure, the total number of correct dockings achieved in all four flexible docking modes. Of the 27 200 dockings 17 922 (65.9%) were in the correct pose. The structures with the best and worst rates of correct docking were 1FVV(c) (86.1%) and 1HCL (32.3%). The average, best, and worst rates of correct docking for the four groups of compounds are, respectively, as follows: group 1, 62.7%, 92.6% (1FVV(a)), and 21.8% (1B38); group 2, 72.5%, 90.1% (1FIN(c)), and 33.0% (1HCL); group 3, 67.5%, 93.5% (1H00), and 30.6% (1HCL); and group 4, 50.4%, 63.2% (1OIT), and 30.6% (1HCL).

Discussion

The reason for doing the work reported herein was to establish if there was one CDK2 structure that, because of its promiscuity, could be used in virtual screening in preference to all others

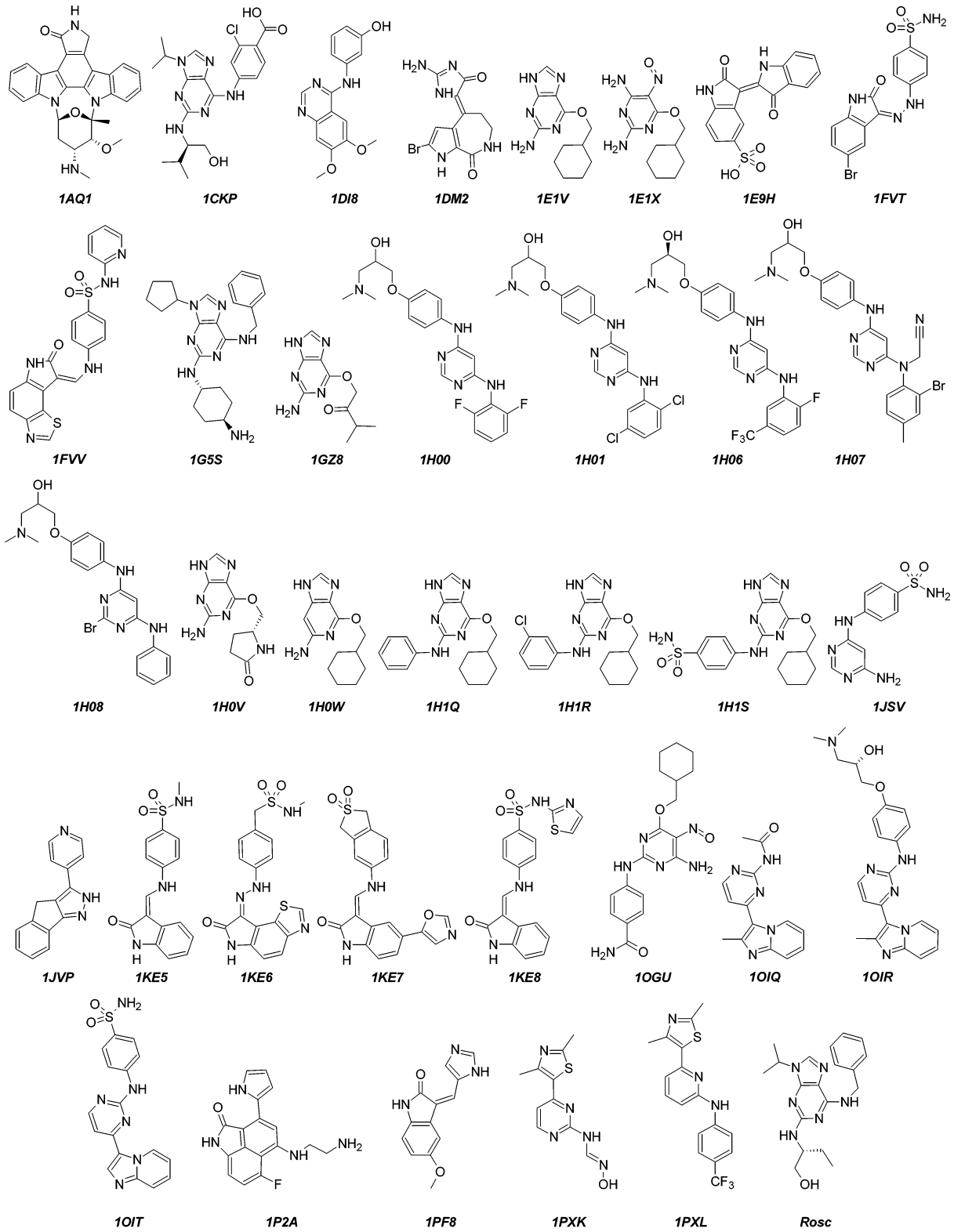

Figure 2. The structures of the 36 ligands of group 4, identified by their CDK2 crystal structure PDB identifiers. References: *1AQ1*;[35] *1CKP*;[50] *1DI8*;[37] *1DM2*;[38] *1E1V, 1E1X*;[39] *1E9H*;[51] *1FVT, 1FVV*;[41] *1G5S*;[42] *1GZ8*;[43] *1H00, 1H01, 1H06, 1H07, 1H08*;[46] *1H0V, 1H0W*;[43] *1H1Q, 1H1R, 1H1S*;[45] *1JSV*;[52] *1JVP*;[48] *1KE5, 1KE6, 1KE7, 1KE8*;[53] *1OGU*;[54] *1OIQ, 1OIR, 1OIT*;[49] *1P2A*;[55] *1PF8*;[56] *1PXK, 1PXL*;[29] *Rosc*.[33]

and to identify those factors that made this structure the most suitable. The means of determining which protein structure might be most suitable for virtual screening was by visual assessment of whether the docked pose of 340 known CDK2 inhibitors replicated the pose observed crystallographically. When screening any database of small molecules, the assessment of how well a ligand binds is based on a visual analysis of the binding pose, together with due consideration of calculated docking scores. As detailed below, many of those ligands that bound incorrectly did so by rotating through 180°: if this had been a screen of an unbiased compound database, many of these compounds could well have been chosen for biochemical assay because their binding pose, although not reproducing the crystallographically observed pose, looks reasonable.

To compare the docking results from different proteins it was necessary to ensure that the docking programs could produce consistent results. Glide dockings were found to be totally reproducible. GOLD dockings were more varied but the degree of variation, as shown in Tables 1 and 2, was not thought to be so great as to preclude the use of GOLD in this study.

Tables 3 and 4 show the docking results for Glide and GOLD, respectively, and Table 5 summarizes these results. The results, the major points of which are highlighted above, across both programs and all docking modes, broadly agree: those protein structures that have a large number of correctly docked ligands with one program operating in one docking mode tend to have a large number of correctly docked ligands with both programs and all docking modes. In all the protein structures, ligand groups, and both programs in both flexible docking modes, ligands with the correct pose tended to have better docking scores than ligands with an incorrect pose, but there was no correlation between docking score and IC_{50}, regardless of all the ligands or only those correctly bound are considered (results not shown). This lack of correlation between docking score and IC_{50} is, at least for GOLD, as expected.[4] The results above are discussed below, first in terms of the ligands, and then in terms of the protein structures.

Group 1 Compounds (Figure 1A). The crystal structures of a few of these ligands show that they dock with the aniline NH and one of the pyrimidine nitrogens, forming hydrogen bonds to the Leu83 carbonyl and NH, respectively[29] (and unpublished data). The range of substitutions on the aniline and thiazole rings should have little affect on the ability of all the group 1 compounds to bind in a similar mode: the substituents can extend into the phosphate binding site or the surrounding solvent. Despite this, across both programs and all four flexible docking modes, only 62.7% of the ligands docked with the correct pose. The most common incorrect binding pose occurred with those ligands where the substitution on the thiazole ring, at R^1 (Figure 1A), was NHR: the thiazole nitrogen and the NH moiety of the R^1 substitution formed the hydrogen bonds to Leu83 NH and CO, respectively. Other incorrect binding modes commonly involved amide or sulfonamide moieties in one of the R groups forming the hydrogen bonds to Leu83.

Group 2 Compounds (Figure 1B−E). Crystallographic data (in-house, unpublished) show that the group 2 compounds bind in a manner similar to the group 1 compounds. Across both programs and all four flexible docking modes, 72.5% of the ligands docked with the correct pose. The reason for the improvement relative to the group 1 compounds is due to the inability of most of the group 2 compounds to dock in the most common incorrect pose adopted by the group 1 compounds. Incorrect binding modes commonly involved amide or sulfonamide moieties in the R groups forming the hydrogen bonds to Leu83.

Group 3 Compounds (Figure 1F). These compounds were all expected to bind in the pose portrayed in, for example, the 1CKP, 1G5S, and roscovitine structures. Across both programs and all four flexible docking modes, 67.5% of the ligands docked with the correct pose. The reason for this success may lie in the fact that these ligands do not penetrate very deeply into the binding site (particularly the phosphate binding site) and have the phenyl ring linked to the purine 6-position oriented out of the pocket.

Group 4 Compounds (Figure 2). These were expected to bind in the pose portrayed in their crystal structures. Across both programs and all four flexible docking modes, 50.4% of the ligands docked with the correct pose. These summary results can be broken down by the structural class of the ligand. In the analysis below, ligands are identified by the name of the PDB structure from which they were extracted. To differentiate them from protein structures of the same name, the ligands are shown in italics.

In the crystal structures, the oxindoles bind to CDK2 with the oxindole NH and carbonyl forming hydrogen bonds to the Glu81 backbone carbonyl and the Leu83 NH, respectively, with an NH group linked to the oxindole 3-position via one (*1E9H*, *1FVT*, *1FVV*, *1KE5*, *1KE6*, *1KE7*, and *1KE8*) or two (*1PF8*) carbon atoms forming a hydrogen bond to the Leu83 carbonyl. Of the 80 dockings of each of these ligands (once into each of 20 protein structures using two programs each in two flexible docking modes), all except *1E9H* (29 out of 80) and *1PF8* (1 out of 80) docked correctly between 46 and 62 times. The reason *1PF8* docks correctly so rarely (compared to the other oxindoles in this study) may be due to its smaller size, permitting greater flexibility in its positioning in the binding site, and the distance between the oxindole ring and the hydrogen bond donor to the Leu83 carbonyl: *1PF8* tends to dock rotated through 180° with the oxindole forming both hydrogen bonds to Leu83 and the imidazole NH forming the hydrogen bond to the Glu81 backbone carbonyl. One possible explanation for the poor rate of correct docking of *1E9H* is the steric hindrance between the 2-oxindole and 3-oxindole rings limiting the ability of the programs to find ligand conformations suitable for good positioning in the varied ATP binding sites. For all these ligands the most frequent incorrect binding mode had the oxindole forming both hydrogen bonds to Leu83.

The crystal structure binding mode of the purines falls into two types. *1CKP*, *1G5S*, and *Rosc* have the same binding mode and, as a class, adopt that pose most successfully, with 63, 59, and 65 out of 80 correct poses, respectively. A possible reason for the success of these dockings is given under the discussion above of the group 3 compounds. *1E1V*, *1GZ8*, *1H0V*, *1H0W*, *1H1Q*, *1H1R*, and *1H1S* bind with the NH and N at the purine 9- and 3-positions, forming hydrogen bonds to the Glu81 carbonyl and Leu83 NH, respectively, and the NH substituent at the purine 2-position forming the hydrogen bond to the Leu83 carbonyl. These compounds bind poorly with between 20 and 49 out of 80 correct poses. Frequently, these ligands dock rotated through 180°, so that the NH and N at the purine 9- and 3-positions form the hydrogen bonds to Leu83. Although not a purine, *1E1X* is structurally related to *1E1V*, binds in a very similar pose (in the crystal structure), and for the same reason docks poorly (20 out of 80).

1H00, *1H01*, *1H06*, *1H07*, *1H08*, *1JSV*, *1OGU*, *1OIR*, *1OIT*, and *1PXL* all have a pyrimidine linked to an aniline with a pyrimidine nitrogen and the aniline NH forming the hydrogen bonds to the Leu83 NH and carbonyl, respectively. *1OIQ* and *1PXK* do not have the aniline ring but do have an NH group attached to the same position on the pyrimidine ring and, in their crystal structures, bind with a similar pose. All these ligands dock correctly with success rates between 30 and 63 out of 80, with most falling in the range 40−50 out of 80: perhaps significantly, given their smaller size and lack of alternative ways of forming the hydrogen bonds, it is *1OIQ* and *1PXK* that have the largest number of correct dockings. Most of the incorrect binding poses involve sulfonamide or amide moieties forming the hydrogen bonds.

The remaining five compounds are structurally dissimilar. *1DM2* binds with 55 out of 80 correct poses: incorrect poses commonly involve the ligand docked with a 180° rotation. Only 19 correct poses of *1P2A* were found, with the incorrect poses

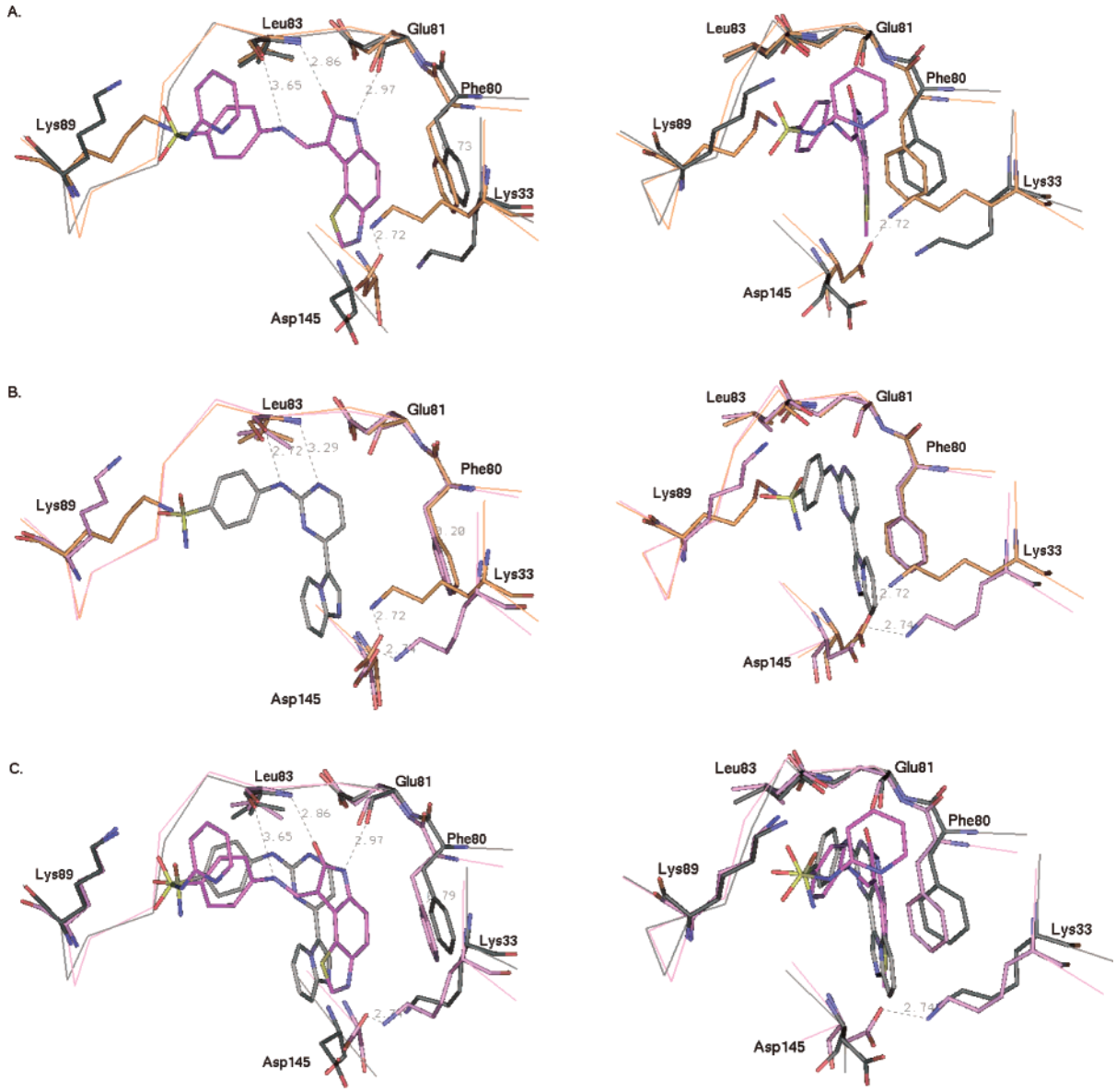

Figure 3. Two views of overlays of each of three pairs of structures. In all the panels, the 1FVV(c) protein carbon atoms are in black and ligand carbon atoms in purple. The 1HCL carbon atoms are in brown. The 1OIT protein carbon atoms are in pink and ligand carbon atoms in gray. The CA trace of some of the residues forming the ATP binding sites are shown, along with selected side chains. In the left-hand half of each panel hydrogen bonds, with distances, between one of the proteins and its ligand are shown, as is the distance between the Phe80 rings and, for 1OIT and 1HCL in both halves of each panel, the distance of the hydrogen bond between Lys33 NZ and Asp145 OD. Panel A, 1FVV(c) and 1HCL; panel B, 1OIT and 1HCL; panel C, 1FVV(c) and 1OIT.

again being due mainly to a 180° rotation of the ligand in the binding site. *1AQ1*, *1DI8*, and *1JVP* do not, in their crystal structures, form the hydrogen bonds required by the docking restraints, though because they have appropriately positioned hydrogen bond donors and acceptors, all could potentially form such hydrogen bonds. For these three ligands only 9, 2, and 1 poses, respectively, were found that reproduced the crystallographically observed pose. Incorrect poses were highly varied but included dockings where the requirements of the hydrogen bonding restraints were met.

Protein Structures. The analysis, below, of the protein structures starts by comparing four pairs of structures for which different rates of correct pose reproduction were achieved and then generalizes the conclusions to all 20 structures.

Across all four flexible docking modes the structures with the best and worst rates of correct docking were 1FVV(c) (86.1%) and 1HCL (32.3%), respectively. These structures are very different: 1HCL is an apo structure while 1FVV(c) is a holo structure with cyclin A. This is reflected in the configuration of the binding site (Figure 3A). In both structures the side chains of Asp145 and of Lys33 project into the binding site: in the 1HCL structure these form a salt bridge (Asp145 OD−Lys33 NZ, 2.72Å) across the binding site—the equivalent distance in 1FVV(c) is 4.05Å with the orientation of the side chains not being conducive to such an electrostatic interaction. In 1HCL these side chains are much higher in the pocket: the distances from Leu83 CA to the Asp145 CG and the Lys33 NZ in 1HCL are 12.31 and 10.33 Å, respectively, and in 1FVV(c) 13.74 and 12.83 Å. This has obvious consequences on the volume into which ligands can dock, though, as discussed in the next two paragraphs, it is probably the position of the Lys33 side chain rather than that of Asp145 that has the greater influence on the rate of correct docking. The position of the Phe80 phenyl ring differs by ~0.7 Å, resulting in the ATP binding site in the 1FVV(c) structure being slightly larger than that of the 1HCL structure. In 1FVV(c) most of the side chain

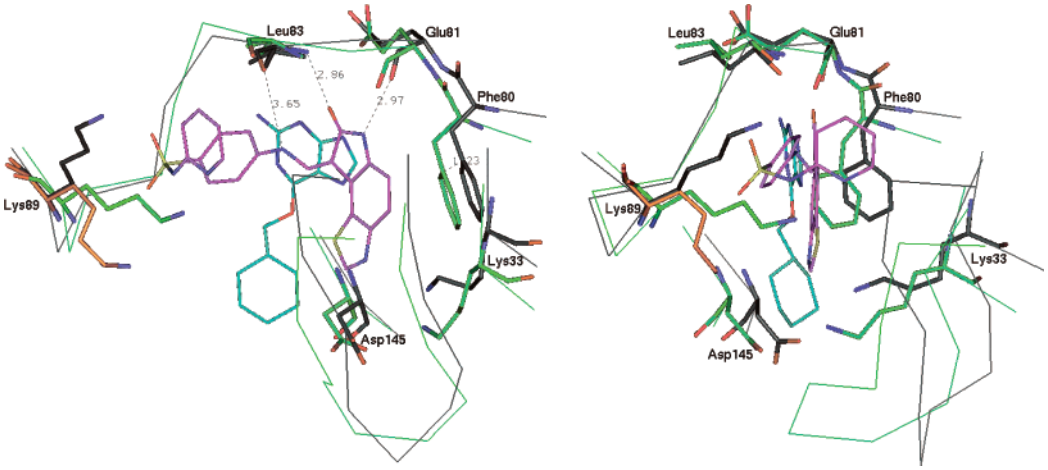

Figure 4. Two views of an overlay of the 1FVV(c) and 1H1P structures. The 1FVV(c) protein carbon atoms are in black and ligand carbon atoms in purple. The 1H1P protein carbon atoms are in green and ligand carbon atoms in pale blue. In brown is Lys89 of 1FIN(c). The CA trace of some of the residues forming the ATP binding sites is shown, along with selected side chains. The pictures in this figure differ from those in Figure 3 in showing the CA trace of residues 9−19 that in 1H1P form the base of the ATP binding site. In the left picture hydrogen bonds, with distances, between one of the proteins and its ligand are shown, as is the distance between the Phe80 rings.

of Tyr15 is missing: ligands docked into 1FVV(c) thus may miss out on some potential contacts but do have a larger volume into which to dock, though the Lys33 and Asp145 side chains probably limit access to this part of the pocket. Tyr15 is part of a loop (residues 9−19) that forms the base of the pocket, but the position of which is quite varied: this is discussed further below.

Of the structures without cyclin A, it was with 1OIT that the highest rate of correct docking (81.0%) was achieved. When this structure is compared with that of 1HCL (32.3%), it can be seen that the positions of the Lys33 and Lys89 side chains are very different in the two structures, with both of them intruding into the ligand binding site to a much greater degree in 1HCL than in 1OIT, thus limiting the accessible volume into which a ligand can dock (Figure 3B). The other difference is that the T-loop of 1OIT is missing. The positions of the Phe80 and Asp145 side chains are very similar. The 1DI8 and 1G5S structures are both structures determined in the absence of cyclin, have a large number of ligands docked in the correct pose, and, as in the 1OIT structure, the Lys33 side chain is oriented away from Phe80. We have previously shown that ligand binding induces significant movement of Lys33 away from Phe80 in CDK2 structures and also results in T-loop disorder.[29] This more open conformation results in increased docking success, and it contrasts with other monomeric structures, e.g. 1GZ8 and 1B38, in which Lys33 intrudes into the binding site.

Given that the number of ligands correctly docked into 1FVV(c) (86.1%; holo structure with cyclin A) and 1OIT (81.0%; the best of the holo structures without cyclin A) was quite similar, the influence of cyclin A on the structure of the binding site was expected to be small, an expectation borne out by an examination of the structures. The positions of the Lys89 side chains are almost identical, and the Lys33 side chains are in similar positions (Figure 3C). The positions of the Phe80 side chains differ by ∼0.8 Å. Probably the biggest difference is in the position of the Asp145 side chain, which is oriented more toward the ATP pocket in 1OIT and is close enough to the Lys33 side chain to form a 2.74 Å salt bridge. That the position of the Asp145 side chain is the same in 1OIT as it is in 1HCL, but that 1OIT has a rate of correct pose reproduction similar to 1FVV(c), suggests that the position of the Asp145 side chain is less critical to correct dockings than the position of the Lys33 side chain. (This observation does not, however, diminish the role of Asp145 in CDK2 inhibition as, in several contexts, it has been shown to contribute significantly to ligand affinity.)

Four of the five CDK2 structures with cyclin A (representing two ligands crystallized as CDK2/cyclin A dimers) are among the top six in terms of the number of correct poses found. The 1FIN structures (1FIN(a), 78.0% and 1FIN(c), 82.2%) are holo structures with ATP, while the 1FVV structures (1FVV(a), 85.0% and 1FVV(c), 86.1%) are holo structures with inhibitors. The structures of these four ATP binding sites are very similar, particularly with regard to the Lys33, Phe80, and Asp145 side chains. The position of the Lys89 side chain does differ between the 1FVV structures on one hand and the 1FIN structures on the other, but in none of the structures does it project into the ATP binding site as it does, for example, in 1HCL (Figure 3A). The fifth CDK2 structure with cyclin A, 1H1P, is a holo structure with inhibitor, but it differs from the other CDK2/cyclin A structures in being phosphorylated on CDK2 Thr160. In terms of the number of correct poses found, it falls in the bottom four (46.0%). As shown in Figure 4, the binding sites of 1FVV(c) and 1H1P differ quite substantially, with the side chain of Lys89 projecting into the ATP binding site and backbone movements affecting the position of the Lys33, Phe80, and Asp145 side chains. Perhaps most significant is the change in the position of the glycine-rich loop composed of residues 9−19, part of which forms the base of the ATP binding site in 1H1P but is more remote in 1FVV(c), which has the open conformation. These factors make the size of the ATP binding site in 1H1P smaller than that in 1FVV(c).

One aspect of the CDK2 structures not yet discussed is the position of the T-loop. In the eight structures that reproduce the greatest number of poses correctly, the T-loop is in either the active conformation, i.e., oriented away from the binding site [1FIN(a), 1FIN(c), 1FVV(a) and 1FVV(c), all structures with cyclin A], or is missing from the PDB file; i.e., it is disordered (1DI8, 1G5S, 1H00, and 1OIT). Five of the six structures with the lowest rate of correct pose reproduction have the T-loop in the inactive conformation, i.e., closed up underneath the loop composed of residues 9−19 (1B38, 1E1V, 1GZ8, 1HCK, and 1HCL): the exception is 1H1P, a structure with cyclin A, in which the T-loop is phosphorylated and in an active conformation, though quite different from the other structures with cyclin A. The position of the T-loop is, however, likely to

be of little direct relevance to the dockings, as it falls outside the defined binding site into which the ligands were docked and reflects other changes in the protein structure.

The points highlighted in the pairwise comparisons above, and shown in Figures 3 and 4, can be applied broadly to all 20 protein structures. When these are compared, the more the side chains of Lys33, Phe80, Lys89, and Asp145 protrude into the ATP binding site and the closer the glycine-rich loop packs in to the base of the ATP binding site, the smaller the number of correctly docked ligands tended to be. These results indicate that some CDK2 structures are better than others, which suggests that these structures should be used in high-throughput docking calculations aimed at identifying new inhibitor chemotypes. Also, certain CDK2 structures have compact or obstructed ATP binding sites, suggesting that their use in virtual screening should be limited. The observation that active CDK2 (cyclin bound) structures give the best rate of pose reproduction is consistent with a concurrent study, performed in this laboratory, that indicates that significant differences in binding mode and energetics are observed for inhibitor structures bound to inactive and active CDK2 (Kontopidis, G. et al. *Chem. Biol.*, in press).

As described above, minimization of the hydrogens and the side chains of the 1HCL and 1FIN(c) structures was performed to determine the affect of these procedures on the number of ligands correctly docked by Glide. Both minimization procedures had only slight affects on the number of ligands correctly docked, and there was no correlation between the level of minimization and the number of ligands correctly docked.

From the data presented in Tables 3−5, a comparison could be made of the relative rates of success of Glide and GOLD in finding correct ligand poses. Such comparisons are, in the opinion of the authors, of questionable value. The binding sites are defined differently, with different sizes, shapes, and centers: the GOLD dockings took place into a sphere of 10 Å radius and a volume of 3053 Å^3, while the Glide dockings took place into a cube with sides of length 14 Å and a volume of 2744 Å^3, though given the hydrogen-bonding restraints and the narrowness of the binding sites the effective difference in the volume into which the ligands were docked is much smaller than the 309 Å^3 implied above. In both programs many parameters were left with their default settings: it is likely that success rates could have been improved by optimizing the parameters. It has been reported by the developers of GOLD that when they repeat GOLD evaluations carried out by other parties the GOLD developers are often able to get better results, presumably because they have more experience with the software (Robin Taylor, personal communication). Furthermore, the nature of the target and the structure of the ligands will influence how successful a program is at docking the ligands into the target: the docking programs differ, and a program that is good for docking into one protein is not necessarily good for docking into another protein (see, for example, refs 10 and 11).

Comparing the results for the fast (Glide SP and GOLD LS) and slow (Glide XP and GOLD SD) docking speeds reveals that there is no relationship between the two. Overall, for both programs the number of ligands correctly docked in the slow mode exceeds the number correctly docked in the fast mode, but this is not true for every protein structure. Presumably, the reason for this difference lies in the details of how the programs work. For GOLD, the number of compounds correctly docked in both modes cannot exceed the smaller of the number of compounds docked in the fast and slow modes and, across all four groups of compounds, is actually less than this [but see the group 3 compounds (Table 4), where this is true for only two of the 20 protein structures]. If the two docking speeds are used in a screening cascade to filter out successively more compounds from a database, then it is a reasonable expectation that the fast docking speed should dock more ligands correctly than the slow docking mode. However, the slow docking speed produces more correct poses: it is reasonable to expect this to be so because the program spends longer docking each ligand. This brings into question the value of a virtual screening cascade.

In summary, although it is not possible to say that any one CDK2 structure is the best for virtual screening, there are some structures that are clearly better than others. The main determinants of this are the volume of the binding site into which the ligands are docked, and the exact orientation of the residues forming the binding site. Given that McGovern and Shoichet[28] reached the same conclusions in their study of 10 other proteins, it is likely that these conclusions will be applicable to most protein structures used in high-throughput docking where the ligand binding site undergoes a conformational rearrangement upon ligand binding. Furthermore, the work reported herein confirms that of Birch et al.,[26] who concluded that a single protein structure could be used in virtual screening but that care must be taken to identify the protein structure that performs best against a wide variety of ligands.

Experimental Section

Protein Structures. The following CDK2 protein structures, taken from the Protein Data Bank,[32] were used in this work: 1AQ1, 1B38, 1DI8, 1DM2, 1E1V, 1E1X, 1FIN, 1FVT, 1FVV, 1G5S, 1GZ8, 1H00, 1H1P, 1HCK, 1HCL, 1JVP, and 1OIT. The 1FIN, 1FVV, and 1H1P structures each contain two CDK2 molecules in the crystallographic asymmetric unit. Only the chain "A" structure from the 1H1P file but both structures from the 1FIN and 1FVV files were used in this work, with the CDK2 protein chain identifiers in the PDB files being used below to differentiate the 1FIN and 1FVV structures: 1FIN(a), 1FIN(c), 1FVV(a), and 1FVV(c). This gives, with the addition of the CDK2−roscovitine structure (a kind gift from Prof. Kim[33]), a total of 20 CDK2 structures. These 20 structures were chosen because of their diversity in terms of resolution, size and shape of the ligand (if any) in the binding site, the presence or absence of other proteins, phosphorylation status, and ATP binding site conformation. Details of the structures are presented in Table 6. There were 18 holo structures (15 with inhibitors and three with ATP in the ATP binding pocket) and two apo structures, with five of the structures being complexes with cyclin A. To enable comparison of the binding site structures, the CA atoms of residues 80−89 were superimposed using the Homology module of InsightII (Accelrys, San Diego, CA). (These residues act as the hinge between the two domains, so they remain relatively static compared to the N- and C-terminal domains.)

Ligands. A total of 340 ATP-competitive inhibitor ligands, with IC_{50} values in the approximate range from 1 nM to 120 μM, were used. The structures were built using the Builder module of InsightII (Accelrys, San Diego, CA) and minimized using the CVFF force field. Conformers were generated in Catalyst (version 4.7; Accelrys, San Diego, CA) using the "best conformer" and default parameters: the 340 compounds yielded 38 684 conformers, an average of ∼114 conformers per compound. These compounds were divided into four groups.

Group 1. This group contained 132 compounds (14 508 conformers, ∼110 conformers per compound) with the 2-anilino-4-(thiazole-5-yl)pyrimidine core structure from our own original CDK inhibitor pharmacophore shown in Figure 1A.

Group 2. This is a structurally more diverse set of 141 compounds (15 323 conformers, ∼109 conformers per compound) from our own extended kinase inhibitor 2-anilino-4-[(hetero)aryl]-pyrimidine pharmacophore with the core structures shown in Figure 1B−E.

Table 6. Protein Structures Used in This Work[a]

PDB code	resolution (Å)	T-loop status	Phos status	missing residues	other molecules	inhibitor	IC_{50} (μM)	ref
1AQ1	2.00	disordered		36−43, 149−161		staurosporine	0.007	35
1B38	2.00	ordered		36−43				36
1DI8	2.20	disordered		37−42, 153−161		4-[3-hydroxyanilino]-6,7-dimethoxyquinazoline	1	37
1DM2	2.10	disordered		36−44, 149−163	ethan-1,2-diol	hymenaldisine	0.04	38
1E1V	1.95	ordered		36−43		NU2058	12	39
1E1X	1.85	ordered		36−43		NU6027	7	39
1FIN	2.30	ordered			cyclin A, ATP			40
1FVT	2.20	disordered		37−43, 154−162		oxindole	0.01	41
1FVV	2.80	ordered			cyclin A	oxindole	0.01	41
1G5S	2.61	disordered		37−45, 14−162		H717	0.048	42
1GZ8	1.30	ordered		37−43		2-amino-6-(3-methyl-2-oxo)butoxypurine	100	43
1H00	1.60	disordered		13, 14, 36−43, 152−161		disubstituted 4,6-bisanilinopyrimidine CDK4 inhibitor	38	44
1H1P	2.10	ordered	T160 P	297−298	cyclin A	NU2058	12	45
1HCK	1.90	ordered		37−40	ATP			46
1HCL	1.80	ordered		37-40				47
1JVP	1.53	disordered		37−43, 153−162		Pkf049-365	1.6	48
1OIT	1.60	disordered		1, 2, 26, 36−46, 152−162		4-[(4-imidazo[1,2-a]pyridine-3-ylpyrimidin-2-yl)amino]benzenesulfonamide	0.003	49
Rosc	2.40	ordered		36−47		roscovitine	0.45	33

[a] In addition to the inhibitors and other molecules shown in the table, some structures have metal ions or phosphates, and some have water molecules. The 1FVT and 1FVV structures both have two CDK2 molecules in the PDB file; both molecules in both structures were used in this work. The 1H1P structure has two CDK2 molecules in the PDB file; only the A chain was used in this work.

Group 3. Purines, the core structure of which is shown in Figure 1F, form group 3. In this group there are 31 compounds with 5620 conformers (∼181 conformers per compound).

Group 4. The final group comprises 36 known CDK2 inhibitors taken from the crystal structures of the following CDK2−inhibitor complexes: 1AQ1, 1CKP, 1DI8, 1DM2, 1E1V, 1E1X, 1E9H, 1FVT, 1FVV, 1G5S, 1GZ8, 1H00, 1H01, 1H06, 1H07, 1H08, 1H0V, 1H0W, 1H1Q, 1H1R, 1H1S, 1JSV, 1JVP, 1KE5, 1KE6, 1KE7, 1KE8, 1OGU, 1OIQ, 1OIR, 1OIT, 1P2A, 1PF8, 1PXK, 1PXL, and roscovitine (hereafter referred to as 'Rosc'). This group had 3233 conformers (∼90 conformers per compound). The structures of these inhibitors, and references for them, are shown in Figure 2. In the remainder of this paper these ligands are identified by the name of the PDB file from which they were extracted: to differentiate them from protein structures of the same name the ligands are shown in italics.

Docking. Ligands were docked into the 20 protein structures using the programs Glide [running under Maestro version 6.0107 and Impact version 2.7112 (Schrodinger Inc., Portland, OR)] and GOLD (version 2.1),[4] according to the protocols described below. With both programs, dockings were performed with the restraint that docked ligands should form hydrogen bonds to Leu83 CO and Leu83 NH. A compound was regarded as having been docked correctly if, by visual inspection, it formed the expected hydrogen bonds, e.g. a group 1 compound should form hydrogen bonds involving a pyrimidine nitrogen and aniline NH.[29] Ligands thus docked should be in approximately the correct position though, because the protein has either no permitted movement (Glide) or only very limited movement (GOLD: rotation of terminal hydrogens to optimize hydrogen bonding), the ligands may not adopt the exact conformations observed crystallographically. For those ligands for which a CDK2−ligand crystal structure was not available, the pose was adjudged to be correct if the interactions between the protein and ligand were the same as those observed in the crystal structures of structurally similar ligands in the CDK2 ATP binding site. The number of ligands correctly docked in each protein structure was recorded. Three ligands among the final group of 36 CDK2 inhibitors do not, in the crystal structures, meet the restraint requirements. These ligands (*1AQ1*, *1DI8*, and *1JVP*) were regarded as correctly docked if, by visual inspection of the docked poses, the same interactions between the ligand and the protein were made as can be observed in the crystal structure. Only the top-ranked pose of each ligand was evaluated. The assessment of the accuracy of a docked pose by the interactions it makes is more subjective than, for example, rmsd from a known structure. Kroemer et al.[34] compared rmsd-based classifications and interaction-based accuracy classifications (IBAC) of docked poses and found significant differences. They concluded that, although more subjective, the IBAC proved to be a more meaningful measure of docking accuracy. A detailed analysis of the scoring functions is beyond the scope of this paper; however, GoldScore (rather than the alternative of ChemScore) was used to score the GOLD-docked poses, and gscore (rather than emodel) was used to score the Glide dockings.

Proteins were prepared for the docking runs by having hydrogens added at pH 7.0 (using the Biopolymer module of InsightII). For use with Glide, all waters, ions, and small molecules other than the ligand were removed from the protein structure, but non-CDK2 proteins were retained. No missing residues, side chains, or non-hydrogen atoms were inserted. Unless otherwise stated, no minimization of the protein structures was performed. Where carried out, minimization was through 1000 cycles of steepest descent using the Discover module of InsightII. The amino and carbonyl groups of Leu83 were marked as being involved in hydrogen bonding. The binding site was identified by placing a 14 Å cube around the ligand. This cube is automatically positioned by the program. Where there was no ligand in the ATP pocket, the protein was superimposed on the 1PXJ structure[29] and the 1PXJ ligand transferred into the empty ATP pocket, with the binding site then being defined around this ligand. For use with GOLD, the protein was treated in exactly the same way except that, additionally, the ligand was removed from the ATP binding site. The binding site was identified by placing a 10 Å radius sphere centered on the position occupied by the ligand non-hydrogen atom with the shortest aggregate distance to the Leu83 amino hydrogen and the Leu83 carbonyl oxygen.

Glide Docking. For the docking runs, all variable parameters had the default values except for the hydrogen-bonding restraints described above, and where noted below. Glide can dock ligands in three modes: (1) rigid docking, (2) standard speed and precision (SP) docking, and (3) extra precision (XP) docking. The 38 684 conformers underwent rigid docking, where the conformation of neither the ligand nor the protein is changed during the docking: the ligand undergoes rigid body translations and rotations in the protein binding site in order to optimally position the ligand in the binding site. In the other two docking modes, the 340 compounds undergo flexible docking (into a static protein) by being additionally permitted to have rotation about single and amide bonds. The extra precision mode is claimed to produce better docked ligand poses than the standard precision mode, but it takes about 10 times longer

than the standard precision mode. For both flexible docking modes, the number of compounds correctly docked in each protein structure, as judged by visual inspection of the docked poses, was recorded. For the rigid docking, there were too many structures to analyze by this method, so only the total number of conformers docked was recorded rather than the number of conformers *correctly* docked—if Glide is unable to find a suitable pose, it does not force a ligand into the binding site but rejects the ligand.

GOLD Docking. For the docking runs, all variable parameters had the default values, except for the hydrogen-bonding restraints described above and where noted below. GOLD has five predefined speeds of operation. The work reported in this paper used only the fastest (library screening (LS) settings) and slowest (standard default (SD) settings) of these five modes. In these two docking modes, the 340 compounds undergo flexible docking (into an almost static protein). The number of compounds correctly docked in each protein structure, as judged by visual inspection of the docked poses, was recorded.

References

(1) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(2) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1−5.

(3) Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput. Aided Mol. Des.* **1996**, *10*, 41−54.

(4) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(5) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289−307.

(6) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.

(7) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem*. **2003**, *46*, 2287−2303.

(8) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks III, C. L. Assessing scoring functions for protein−ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.

(9) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Struct. Funct. Bioinf.* **2004**, *57*, 225−242.

(10) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: Comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.

(11) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Struct. Funct. Bioinf.* **2004**, *56*, 235−249.

(12) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem*. **1999**, *42*, 5100−5109.

(13) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.

(14) Terp, G. E.; Johansen, B. N.; Christensen, I. T.; Jørgensen, F. S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein−ligand binding affinities. *J. Med. Chem.* **2001**, *44*, 2333−2343.

(15) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002**, *20*, 281−295.

(16) Jacobsson, M.; Lidén, P.; Stjernschantz, E.; Boström, H.; Norinder, U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.* **2003**, *46*, 5781−5789.

(17) Joseph-McCarthy, D. Computational approaches to structure-based ligand design. *Pharmacol. Ther*. **1999**, *84*, 179−191.

(18) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol*. **2000**, *57*, 213−218.

(19) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct. Funct. Genet.* **2002**, *47*, 409−443.

(20) Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. Lead discovery using molecular docking. *Curr. Opin. Chem. Biol*. **2002**, *6*, 439−446.

(21) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct*. **2003**, *32*, 335−373.

(22) Alvarez, J. C. High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.* **2004**, *8*, 365−370.

(23) Hou, T.; Xu, X. Recent development and application of virtual screening in drug discovery: An overview. *Curr. Pharm. Des.* **2004**, *10*, 1011−1033.

(24) Oprea, T. I.; Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.* **2004**, *8*, 349−358.

(25) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076−5084.

(26) Birch, L.; Murray, C. W.; Hartshorn, M. J.; Tickle, I. J.; Verdonk, M. L. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Comput. Aided Mol. Des.* **2002**, *16*, 855−869.

(27) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.

(28) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895−2907.

(29) Wu, S. Y.; McNae, I.; Kontopidis, G.; McClue, S. J.; McInnes, C.; Stewart, K. J.; Wang, S.; Zheleva, D. I.; Marriage, H.; Lane, D. P.; Taylor, P.; Fischer, P. M.; Walkinshaw, M. D. Discovery of a novel family of CDK inhibitors with the program LIDAEUS: Structural basis for ligand-induced disordering of the activation loop. *Structure* **2003**, *11*, 399−410.

(30) Wang, S., Meades, C., Wood, G., Osnowski, A., Anderson, S., Yuill, R., Thomas, M., Mezna, M., Jackson, W., Midgley, C., Griffiths, G., Fleming, I., Green, S., McNae, I., Wu, S.-Y., McInnes, C., Zheleva, D., Walkinshaw, M. D., Fischer, P. M. 2-Anilino-4-(thiazol-5-yl)pyrimidine CDK inhibitors: Synthesis, SAR analysis, X-ray crystallography, and biological activity. *J. Med. Chem.* **2004**, *47*, 1662−1675.

(31) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(32) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(33) De Azevedo, W. F.; Leclerc, S.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S.-H. Inhibition of cyclin-dependent kinases by purine analogues: Crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.* **1997**, *243*, 518−526.

(34) Kroemer, R. T., Vulpetti, A., McDonald, J. J., Rohrer, D. C., Trosset, J.-Y., Giordanetto, F., Cotesta, S., McMartin, C., Kihlén, M., Stouten, P. F. Assessment of docking poses: Interactions-based accuracy classification (IBAC) versus crystal structure deviations. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 871−881.

(35) Lawrie, A. M.; Noble, M. E. M.; Tunnah, P.; Brown, N. R.; Johnson, L. N.; Endicott, J. A. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. *Nature Struct. Biol*. **1997**, *4*, 796−801.

(36) Brown, N. R.; Noble, M. E. M.; Lawrie, A. M.; Morris, M. C.; Tunnah, P.; Divita, G.; Johnson, L. N.; Endicott, J. A. Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *J. Biol. Chem.* **1999**, *274*, 8746−8756.

(37) Shewchuk, L.; Hassell, A.; Wisely, B.; Rocque, W.; Holmes, W.; Veal, J.; Kuyper, L. F. Binding mode of the 4-anilinoquinazoline class of protein kinase inhibitor: X-ray crystallographic studies of 4-anilinoquinazolines bound to cyclin-dependent kinase 2 and p38 kinase. *J. Med. Chem.* **2000**, *43*, 133−138.

(38) Meijer, L.; Thunnissen, A.-M. W. H.; White, A. W.; Garnier, M.; Nikolic, M.; Tsai, L.-H.; Walter, J.; Cleverley, K. E.; Salinas, P. C.; Wu, Y.-Z.; Biernat, J.; Mandelkow, E.-M.; Kim, S.-H.; Pettit, G. R. Inhibition of cyclin-dependent kinases, GSK-3β and CK1 by hymenialdisine, a marine sponge constituent. *Chem. Biol.* **2000**, *7*, 51−63.

(39) Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E. M.; Sausville, E. A.; Schultz, R.; Yu, W. Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J. Med. Chem.* **2000**, *43*, 2797−2804.

(40) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massagué, J.; Pavletich, N. P. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **1995**, *376*, 313−320.

(41) Davis, S. T.; Benson, B. G.; Bramson, H. N.; Chapman, D. E.; Dickerson, S. H.; Dold, K. M.; Eberwein, D. J.; Edelstein, M.; Frye, S. V.; Gampe, R. T. Jr.; Griffin, R. J.; Harris, P. A.; Hassell, A. M.; Holmes, W. D.; Hunter, R. N.; Knick, V. B.; Lackey, K.; Lovejoy, B.; Luzzio, M. J.; Murray, D.; Parker, P.; Rocque, W. J.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Prevention of chemotherapy-induced alopecia in rats by CDK inhibitors. *Science* **2001**, *291*, 134−137.

(42) Dreyer, M. K.; Borcherding, D. R.; Dumont, J. A.; Peet, N. P.; Tsay, J. T.; Wright, P. S.; Bitonti, A. J.; Shen, J.; Kim, S.-H. Crystal structure of human cyclin-dependent kinase 2 in complex with the adenine-derived inhibitor H717. *J. Med. Chem.* **2001**, *44*, 524−530.

(43) Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Curtin, N. J.; Davies, T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Whitfield, H. J. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O^6-substituted guanine derivatives. *J. Med. Chem.* **2002**, *45*, 3381−3393.

(44) Beattie, J. F.; Breault, G. A.; Ellston, R. P. A.; Green, S.; Jewsbury, P. J.; Midgley, C. J.; Naven, R. T.; Minshull, C. A.; Pauptit, R. A.; Tucker, J. A.; Pease, J. E. Cyclin-dependent kinase 4 inhibitors as a treatment for cancer. Part 1: Identification and optimisation of substituted 4,6-bis anilino pyrimidines. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2955−2960.

(45) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Wang, L.; Whitfield, H. J. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nature Struct. Biol.* **2002**, *9*, 745−749.

(46) Schulze-Gahmen, U.; De Bondt, H. L.; Kim, S.-H. High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: Bound waters and natural ligand as guides for inhibitor design. *J. Med. Chem.* **1996**, *39*, 4540−4546.

(47) Schulze-Gahmen, U.; Brandsen, J.; Jones, H. D.; Morgan, D. O.; Meijer, L.; Vesely, J.; Kim, S.-H. Multiple modes of ligand recognition: Crystal structures of cyclin-dependent kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins: Struct. Func. Genet.* **1995**, *22*, 378−391.

(48) Furet, P.; Meyer, T.; Strauss, A.; Raccuglia, S.; Rondeau, J.-M. Structure-based design and protein X-ray analysis of a protein kinase inhibitor. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 221−224.

(49) Anderson, M.; Beattie, J. F.; Breault, G. A.; Breed, J.; Byth, K. F.; Culshaw, J. D.; Ellston, R. P. A.; Green, S.; Minshull, C. A.; Norman, R. A.; Pauptit, R. A.; Stanway, J.; Thomas, A. P.; Jewsbury, P. J. Imidazo[1,2-*a*]pyridines: A potent and selective class of cyclin-dependent kinase inhibitors identified through structure-based hybridisation. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3021−3026.

(50) Gray, N. S.; Wodicka, L.; Thunnissen, A.-M. W. H.; Norman, T. C.; Kwon, S.; Espinoza, F. H.; Morgan, D. O.; Barnes, G.; LeClerc, S.; Meijer, L.; Kim, S.-H.; Lockhart, D. J.; Schultz, P. G. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **1998**, *281*, 533−538.

(51) Davies, T. G.; Tunnah, P.; Meijer, L.; Marko, D.; Eisenbrand, G.; Endicott, J. A.; Noble, M. E. M. Inhibitor binding to active and inactive CDK2: The crystal structure of CDK2-cyclin A/indirubin-5-sulphonate. *Structure* **2001**, *9*, 389−397.

(52) Clare, P. M.; Poorman, R. A.; Kelley, L. C.; Watenpaugh, K. D.; Bannow, C. A.; Leach, K. L. The cyclin-dependent kinases cdk2 and cdk5 act by a random, anticooperative kinetic mechanism. *J. Biol. Chem.* **2001**, *276*, 48292−48299.

(53) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T. Jr.; Harris, P. A.; Hassell, A.;, Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): Design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339−4358.

(54) Sayle, K. L.; Bentley, J.; Boyle, F. T.; Calvert, A. H.; Cheng, Y.; Curtin, N. J.; Endicott, J. A.; Golding, B. T.; Hardcastle, I. R.; Jewsbury, P.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Parsons, R. J.; Pratt, D. J.; Wang, L. Z.; Griffin, R. J. Structure-based design of 2-arylamino-4-cyclohexylmethyl-5-nitroso-6-amino-pyrimidine inhibitors of cyclin-dependent kinases 1 and 2. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3079−3082.

(55) Liu, J.-J.; Dermatakis, A.; Lukacs, C.; Konzelmann, F.; Chen, Y.; Kammlott, U.; Depinto, W.; Yang, H.; Yin, X.; Chen, Y.; Schutt, A.; Simcox, M. E.; Luk, K.-C. 3,5,6-trisubstituted naphthostyrils as CDK2 inhibitors. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2465−2468.

(56) Moshinsky, D. J.; Bellamacina, C. R.; Boisvert, D. C.; Huang, P.; Hui, T.; Jancarik, J.; Kim, S.-H.; Rice, A. G. SU9516: Biochemical analysis of cdk inhibition and crystal structure in complex with cdk2. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 1026−1031.

JM050554I