SUPPORTING INFORMATION

DFGmodel: Predicting Protein Kinase Structures in Inactive States for Structure-Based Discovery of Type-II Inhibitors

Peter Man-Un Ung* **and Avner Schlessinger***

Department of Pharmacology and Systems Therapeutics, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States

* **Corresponding Author**s:

Peter Man-Un Ung: peter.ung@mssm.edu
Avner Schlessinger: avner.schlessinger@mssm.edu

SUPPORTING FIGURES

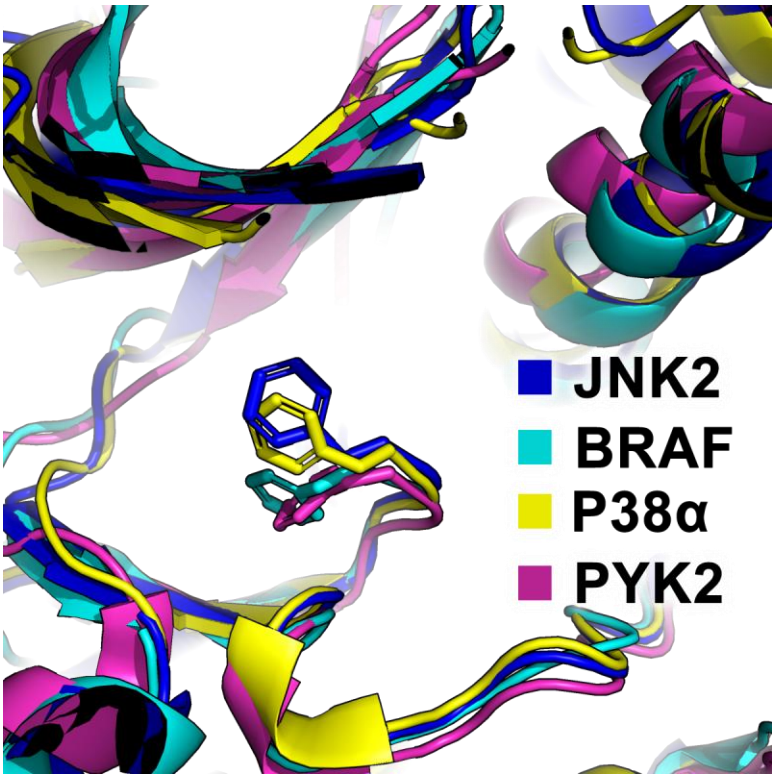

Figure S1. DFG-Phenylalanine side chain adopts multiple conformations. The type-II kinase inhibitor BIRB796 can be found in 4 complex structures: 3NPC (JNK2), 1KV2 (P38α), 3FZS (PYK2), and 4JVG (BRAF). All four complexes share similar ligand binding pose of BIRB796, while the exact position of the DFG-phenylalanine side chain (colored sticks) varies among the structures.

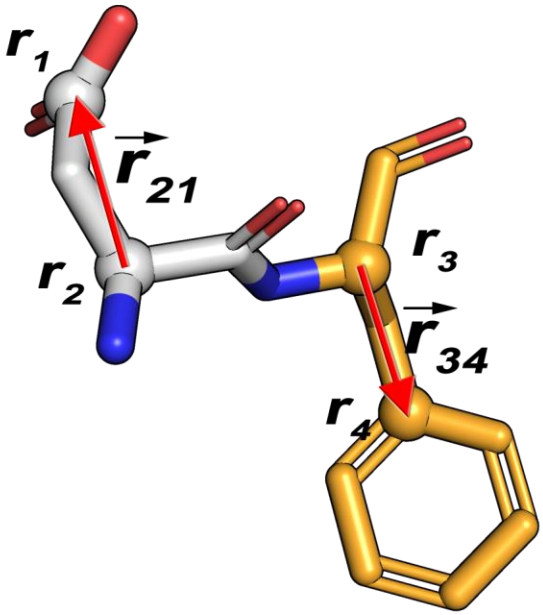

Figure S2. Atomic coordinates and vectors for DFG-motif vector calculations. Residue at position (D), typically DFG-Asp, is colored in white, whereas residue at position (D+1), typically DFG-Phe, is colored in gold. r_1, r_2, r_3, and r_4 are the atomic coordinates of (D):Cγ, (D):Cα, (D+1):Cα, and (D+1):Cγ, respectively.

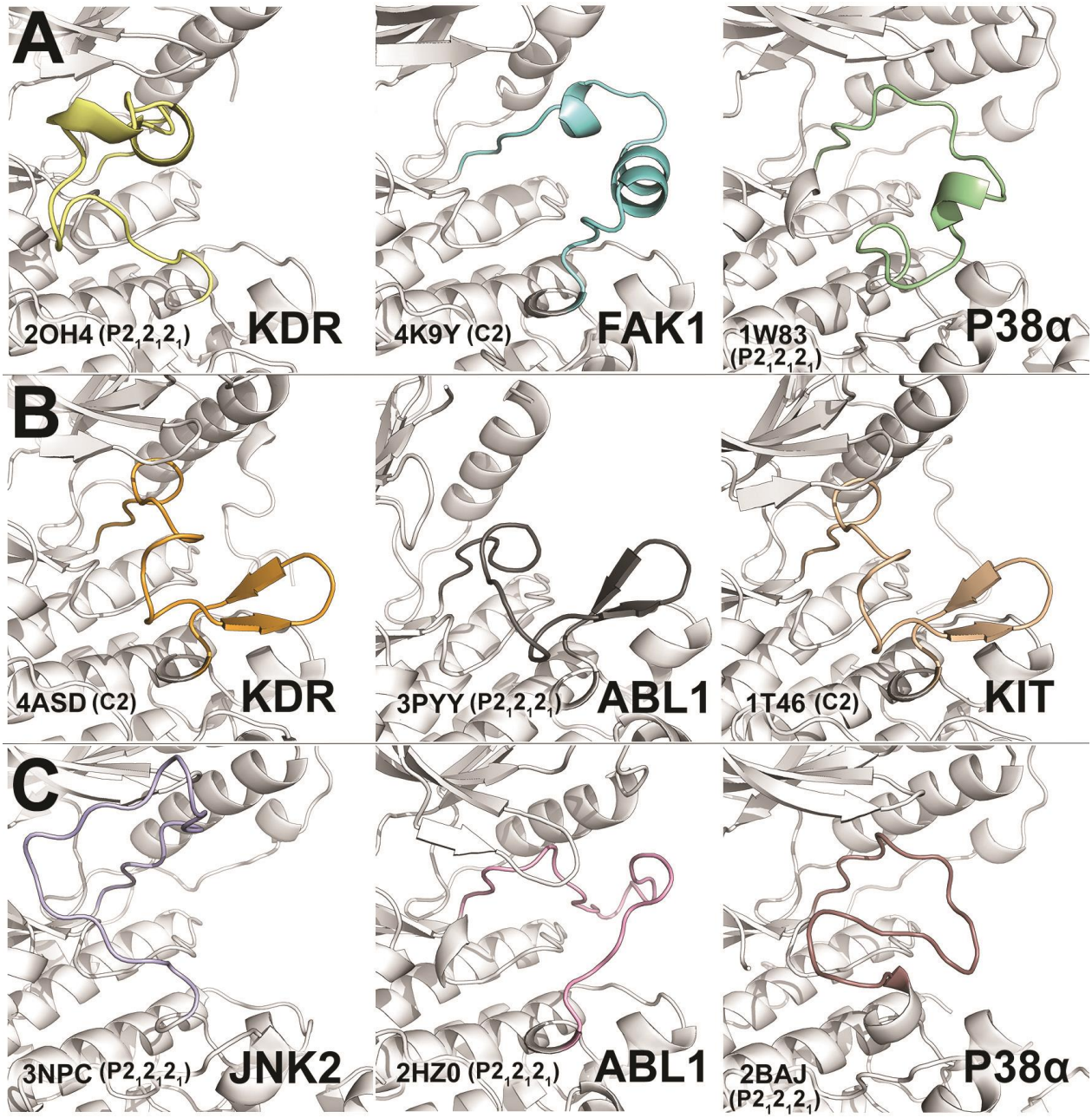

Figure S3. Activation loop in DFG-out kinase structures is structurally diverse. Many kinase structures do not have well-ordered structure of the A-loop beyond the DFG-motif. We excluded DFG-out structures with disordered residue at or before position (D+3) in the A-loop, which results in 60 S/T-kinase and 64 Y-kinase accepted structures. Among these, 13 S/T-kinase and 30 Y-kinase structures have fully resolved A-loop that folds into various secondary structure elements (colored cartoon), such as **(A)** α-helix, **(B)** β-hairpin, or **(C)** loop. These structures belong to different space groups ($P2_12_12_1$, C2, and $P3_221$), suggesting that the different conformations are unlikely related to crystal packing (listed next to the PDB ID). Several kinases have been observed to adopt more than one A-loop conformation, e.g. KDR (2OH4 and 4ASD), ABL1 (3PYY and 2HZ0), and P38α (1W83 and 2BAJ). Of these examples, only KDR's 2OH4 has a phosphorylated Tyr in the A-loop.

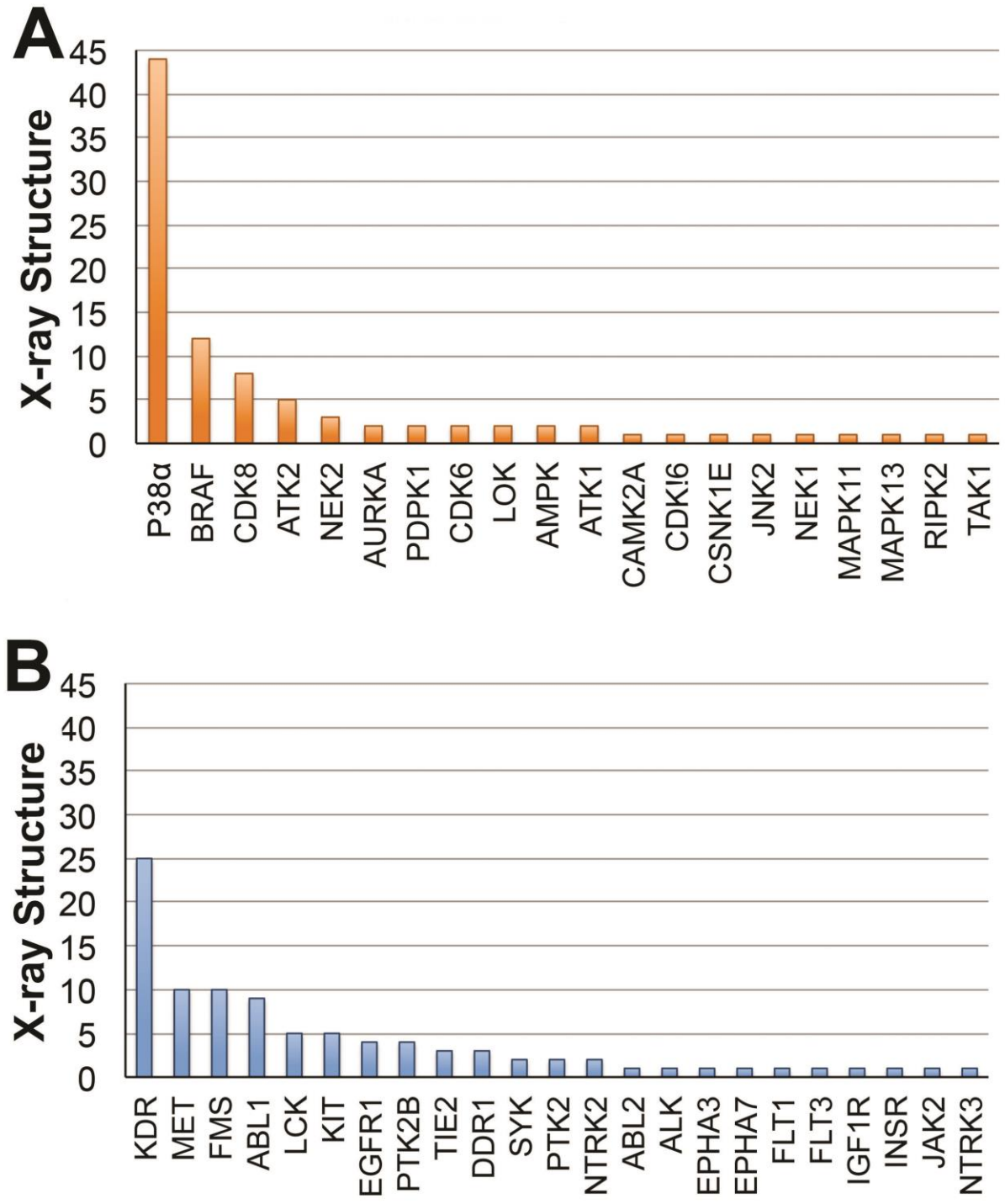

Figure S4. Number of experimentally determined DFG-out structures. **(A)** S/T-kinases, and **(B)** Y-kinases deposited in the PDB.

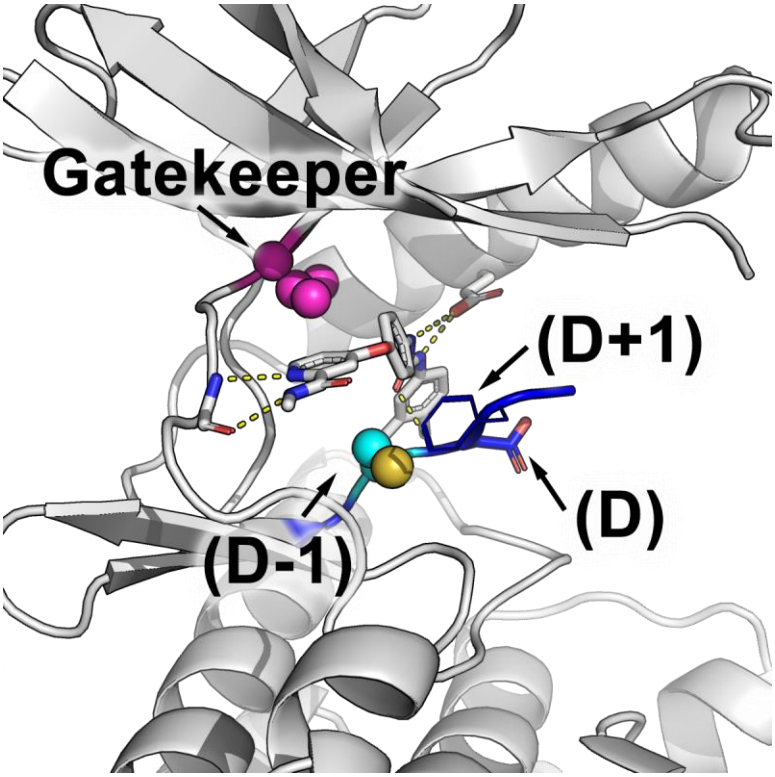

Figure S5. The gatekeeper and (D-1) residues. These two residues (colored spheres) are in close proximity to the type-II kinase inhibitor (sorafenib, white sticks) in KDR (PDB: 4ASD).

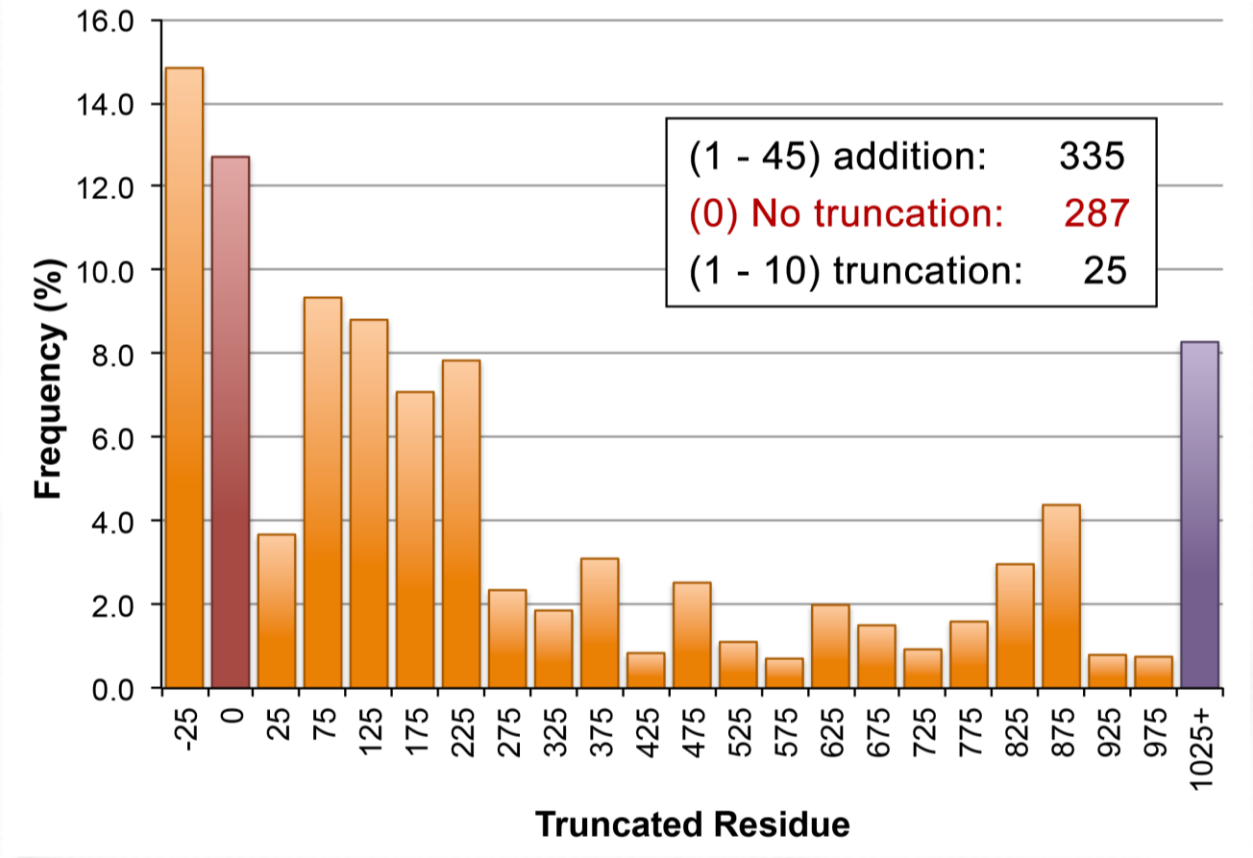

Figure S6. Length distribution of human protein kinase chains. Many protein kinases are multiple-domain proteins that contain regulatory and non-catalytic domains, in addition to the conserved catalytic kinase domain. We surveyed 2,258 chains of human protein kinases found in PDB (excluding atypical kinases) and defined the canonical isoform of the protein found in UniProt as the full-length protein. By comparing the PDB kinase chains to the canonical sequences, 287 (~12.7%) chains were found to have no truncation. Proteins that were expressed with expression tag or had partially-cleaved tag (protein chains with additional residues, e.g. PDB: 4ACC), as well as those that have disordered or slightly truncated terminal residues (<10 residues, e.g. PDB: 3K3I) were defined as “full-length” (647 kinase chains; ~28.7%). 46.4% of protein chains have 50- to 500-residue truncation, which corresponds to the truncation of one or more domains, e.g., SH2-, SH3-, and/or PH-domains. 25.0% chains have longer truncation (>500 residues), most of which are receptor tyrosine kinases (~70% of Y-kinase chains, ~7% of S/T-kinase chains) that include extracellular and transmembrane domains.

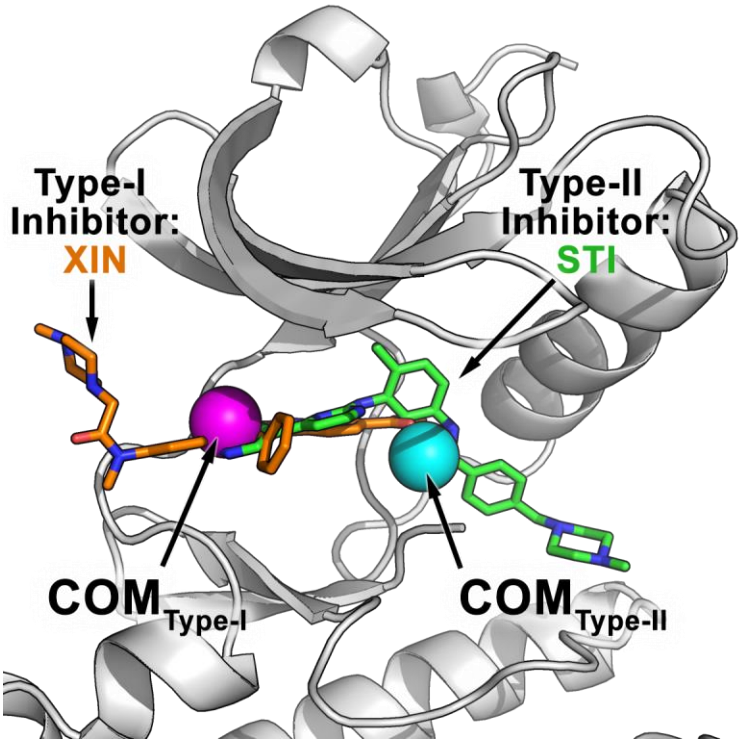

Figure S7. Definition of type-I and type-II kinase inhibitors. 1,458 co-crystals of protein kinase with type-I or type-II kinase inhibitors were obtained from the PDB. The distance between the center-of-mass (COM) of the inhibitor to the COM of a known type-I (ligand 215 from 2FB8 for S/T-kinases and XIN from 3C7Q for Y-kinases) or type-II (imitanib from 3HEC and 2PL0 for S/T- and Y-kinases, respectively) kinase inhibitor was calculated. Ligands with $COM_{type\text{-}I}$ > 5.00 Å and $COM_{type\text{-}II}$ < 3.05 Å are considered type-II ligands (118 ligands), while all others are considered type-I ligands (1340 ligands). In addition to these ligands, analogs of the type-II inhibitor PD5 from 3EL8[1] (AD-36, AD-57, AD-58, AD-80, and AD-81) that have wide anti-kinase spectrum[2] was added to the ligand set.

SUPPORTING TABLES

Table S1. Size-Score of select protein kinases

	Gatekeeper[a] Residue (Size[c])	(D-1)[b] Residue (Size[c])	Size Sum[d]	Channel Width[e]	Type-II Inhibitor[f] (K_d < 500 nM)
ABL1	T (2)	A (1)	3	wide	10
AURKB	T (2)	A (1)	3	wide	4
BLK	T (2)	A (1)	3	wide	5
BRAF	T (2)	G (1)	3	wide	2
CDK8	F (3)	A (1)	4	medium	4
DYRK2	L (3)	I (3)	6	narrow	1
EGFR	T (2)	T (2)	4	medium	2
EPHA3	T (2)	S (1)	3	wide	3
JNK2	M (3)	L (3)	6	narrow	4
KIT	T (2)	C (1)	3	wide	14
LCK	T (2)	A (1)	3	wide	10
LOK	I (3)	A (1)	4	medium	6
LTK	L (3)	G (1)	4	medium	2
MEK2	M (3)	C (1)	4	medium	1
MET	L (3)	A (1)	4	medium	1
P38α	T (2)	L (3)	5	narrow	7
P38γ	M (3)	L (3)	6	narrow	3
PDGFRA	T (2)	C (1)	3	wide	12
RET	V (2)	S (1)	3	wide	10
SRC	T (2)	A (1)	3	wide	4
TAK1	M (3)	C (1)	4	medium	3
TIE1	I (3)	A (1)	4	medium	5
KDR	V (2)	C (1)	3	wide	9

[a] Gatekeeper Residue is the amino acid type in the gatekeeper position of the binding site

[b] (D-1) Residue marks the amino acid type in the position preceding the DFG-motif

[c] Size marks the estimated size of the amino acid type: small residue (G, A, C, S = 1); medium residue (T, V = 2); large residue (I, L, M, F = 3)

[d] Size Sum marks the sum of the Size of the gatekeeper and (D-1) residue

[e] Channel Width is defined by Size Sum: wide (Sum = 3); medium (Sum = 4); narrow (Sum > 5). The Channel Width corresponds to the estimated width of the channel connecting the ATP binding site and the DFG pocket

[f] Type-II Inhibitor represents the number of type-II inhibitors that inhibit the target kinase significantly (K_d < 500 nM) from Davis *et al.*[3] and Deng *et al.*[4] (total 14 inhibitors)

Table S2. Composition of amino acid type at the gatekeeper and (D-1) positions in human kinome

	Gatekeeper Residue				(D-1) Residue		
	All Kinases[a]	DFG-Out[b]	DFG-In[c]		All Kinases[a]	DFG-Out[b]	DFG-In[c]
I	11	4	7	G	69	11	58
L	82	4	78	A	144	14	130
M	184	15	169	I	46	1	45
F	64	7	57	L	15	4	11
T	92	12	80	S	69	4	65
				T	78	4	74
				C	48	3	45

[a] All Kinases marks the number of different amino acid type found in human kinome, excluding atypical kinases

[b] DFG-Out marks the number of different amino acid type found in kinases with DFG-out conformations deposited in PDB

[c] DFG-in marks the number of kinase that have not been co-crystalized in DFG-out conformations

Table S3. 18 templates for DFGmodel

Serine/Threonine-Kinases		Tyrosine-Kinases	
Name	PDB ID	Name	PDB ID
PDK1- AGC (O15530)	3QC4_A	JAK (O60674)	3UGC_A
TAK1 - STE (O43318)	2YIY_A	IR - Insulin (P06213)	3ETA_A
STK10 - STE (O94804)	4EQU_A	LCK – Src (P06239)	2OFV_A, 2OG8_A
BRAF - TKL (P15056)	1UWH_A, 3II5_A, 4JVG_A	MET (P08581)	3F82_A, 4EEV_A
JNK - GMGC (P45984)	3NPC_A*	KDR – PDGF (P35968)	2OH4_A*, 2P2I_A, 2QU5_A, 3EWH_A, 4ASD_A*
CDK8 - CMGC (P49336)	4F7N_A	SYK - SYK (P43405)	3TUB_A
MAPK13 - GMGC (O15264)	4EYJ_A	PTK2 - FAK (Q05397)	4K9Y_A*
P38α – GMGC (Q16539)	2BAJ_A*, 3GCV_A, 3HV5_A, 3KQ7_A*	DDR1 - Insulin (Q08345)	3ZOS_A, 4BKJ_A
LCK – TK/Src (P06239)	2PL0_A*, 2OG8_A	PTK2B – FAK (Q14289)	3FZT_A
KDR – TK/PDGF (P35968)	3EWH_A	EPHA7 - Ephrin (Q15375)	3DKO_A
MET – TK (P08581)	3L8V_A	NTRK2 - Insulin (Q16620)	4AT5_A*
OTK2B – TK/FAK (Q14289)	4H1J_A		

* activation loop is ordered

Table S4. Z-DOPE scores for crystal structures and models

	Crystal Structures		Homology Models			
	DFG-in	DFG-out	DFGmodel	MODBASE		
ABL1[b]	-1.61 ± 0.13	-1.60 ± 0.14	-1.44 ± 0.04	-1.54	DFG-in	1OPK
BRAF[a]	-1.71 ± 0.13	-1.49 ± 0.15	-1.29 ± 0.04	-0.91	DFG-out	1UWH
KIT[b]	-1.42 ± 0.15	-1.80 ± 0.09	-1.07 ± 0.06	-0.25	DFG-in	3G0E
LCK[b]	-1.95 ± 0.05	-1.72 ± 0.09	-1.79 ± 0.06	-1.66	DFG-in	1QCF (Src)
P38α[a]	-1.47 ± 0.08	-1.38 ± 0.07	-1.10 ± 0.03	-0.73	DFG-out	2BAJ
SRC[b]	-1.58 ± 0.23	-1.67 ± 0.10	-1.51 ± 0.05	-1.57	DFG-in	1FMK
EPHA3[b]	-1.95 ± 0.04	-1.96 ± 0.01	-1.76 ± 0.05	-0.43	DFG-in	1JPA (EphB2)
KDR[b]	-1.66 ± 0.16	-1.73 ± 0.12	-1.46 ± 0.05	-1.08	DFG-out	3BE2
CDK8[a]	-1.55 ± 0.01	-1.49 ± 0.08	-1.21 ± 0.07	-0.78	DFG-out	3RGF
JNK2[a]	-1.54 ± 0.01	-1.75 ± 0.01	-1.06 ± 0.06	-0.72	DFG-in	3E7O
Average	-1.66 ± 0.10	-1.61 ± 0.09	-1.33 ± 0.05	-1.08		
Median	-1.63 ± 0.12	-1.64 ± 0.09	-1.31 ± 0.05	-1.17		
StDev	0.17		0.24			

[a] Serine/threonine kinase
[b] Tyrosine kinase

Table S5. Comparison of DFG-in crystal structures

	TM-Score			RMSD (Å)		
	N-lobe	C-lobe	Full domain	N-lobe	C-lobe	Full domain
ABL1[b]	0.91 ± 0.02	0.98 ± 0.02	0.96 ± 0.01	1.01 ± 0.23	0.60 ± 0.35	1.01 ± 0.14
BRAF[a]	0.90 ± 0.05	0.93 ± 0.07	0.93 ± 0.05	1.41 ± 0.59	0.51 ± 0.22	1.55 ± 0.83
KIT[b]	0.81 ± 0.13	0.48 ± 0.15	0.80 ± 0.13	1.74 ± 0.57	0.88 ± 0.37	2.02 ± 0.75
LCK[b]	0.97 ± 0.04	0.99 ± 0.01	0.99 ± 0.01	0.42 ± 0.29	0.20 ± 0.09	0.38 ± 0.23
P38α[a]	0.98 ± 0.01	0.84 ± 0.03	0.97 ± 0.01	0.60 ± 0.15	0.23 ± 0.03	0.71 ± 0.15
SRC[b]	0.87 ± 0.06	0.92 ± 0.02	0.53 ± 0.22	1.45 ± 0.47	1.28 ± 0.40	2.36 ± 0.86
EPHA3[b]	0.96 ± 0.01	0.92 ± 0.04	0.95 ± 0.02	0.46 ± 0.20	0.68 ± 0.47	0.91 ± 0.48
KDR[b]	0.96 ± 0.02	0.80 ± 0.04	0.95 ± 0.02	0.71 ± 0.24	0.28± 0.09	0.84 ± 0.37
CDK8[a]	0.95 ± 0.00	0.86 ± 0.00	0.96 ± 0.00	0.62 ± 0.00	0.57 ± 0.00	0.64 ± 0.00
JNK2[a]	-	-	-	-	-	-
Mean	0.92 ± 0.04	0.88 ± 0.14	0.90 ± 0.13	0.93 ± 0.41	0.53 ± 0.32	1.09 ± 0.57

[a] Serine/threonine kinase
[b] Tyrosine kinase

Table S6. Comparison between DFG-out crystal structures

	TM-Score			RMSD (Å)		
	N-lobe	C-lobe	Full domain	N-lobe	C-lobe	Full domain
ABL1[b]	0.89 ± 0.01	0.91 ± 0.01	0.90 ± 0.01	1.18 ± 0.34	1.49 ± 0.24	2.00 ± 0.15
BRAF[a]	0.94 ± 0.01	0.91 ± 0.01	0.94 ± 0.01	1.16 ± 0.07	0.68 ± 0.28	1.24 ± 0.26
KIT[b]	0.82 ± 0.00	0.60 ± 0.00	0.82 ± 0.00	1.34 ± 0.00	1.11 ± 0.00	1.69 ± 0.00
LCK[b]	0.87 ± 0.05	0.90 ± 0.01	0.89 ± 0.02	1.15 ± 0.59	0.89 ± 0.69	1.47 ± 0.54
P38α[a]	0.97 ± 0.02	0.92 ± 0.05	0.94 ± 0.04	0.55 ± 0.24	0.96 ± 0.38	1.01 ± 0.43
SRC[b]	0.92 ± 0.05	0.96 ± 0.01	0.94 ± 0.03	0.83 ± 0.10	0.63 ± 0.19	0.96 ± 0.11
EPHA3[b]	-	-	-	-	-	-
KDR[b]	0.78 ± 0.05	0.78 ± 0.06	0.78 ± 0.05	0.82 ± 0.49	0.29 ± 0.05	0.82 ± 0.49
CDK8[a]	0.86 ± 0.02	0.83 ± 0.03	0.86 ± 0.02	0.48 ± 0.08	0.24 ± 0.02	0.48 ± 0.08
JNK2[a]	-	-	-	-	-	-
Mean	0.87 ± 0.07	0.88 ± 0.12	0.87 ± 0.07	0.89 ± 0.38	0.70 ± 0.41	1.11 ± 0.54

[a] Serine/threonine kinase
[b] Tyrosine kinase

Table S7. Comparison between DFGmodel models and DFG-out crystal structures

	TM-Score			RMSD (Å)		
	N-lobe	C-lobe	Full Domain	N-lobe	C-lobe	Full domain
ABL1[b]	0.89 ± 0.02	0.90 ± 0.02	0.89 ± 0.01	1.29 ± 0.16	1.03 ± 0.44	1.54 ± 0.35
BRAF[a]	0.94 ± 0.02	0.97 ± 0.01	0.97 ± 0.01	0.93 ± 0.18	0.63 ± 0.22	0.92 ± 0.20
KIT[b]	0.93 ± 0.01	0.98 ± 0.01	0.97 ± 0.01	1.27 ± 0.02	0.70 ± 0.02	1.09 ± 0.01
LCK[b]	0.89 ± 0.04	0.96 ± 0.01	0.94 ± 0.03	1.26 ± 0.09	0.78 ± 0.23	1.30 ± 0.30
P38α[a]	0.97 ± 0.01	0.94 ± 0.04	0.95 ± 0.03	0.61 ± 0.05	0.86 ± 0.17	1.01 ± 0.18
SRC[b]	0.89 ± 0.04	0.97 ± 0.01	0.95 ± 0.02	1.14 ± 0.05	0.80 ± 0.17	1.10 ± 0.11
EPHA3[b]	0.89 ± 0.00	0.98 ± 0.00	0.94 ± 0.00	1.39 ± 0.00	0.67 ± 0.00	1.73 ± 0.00
KDR[b]	0.93 ± 0.02	0.97 ± 0.01	0.96 ± 0.01	0.87 ± 0.33	0.73 ± 0.18	1.00 ± 0.15
CDK8[a]	0.80 ± 0.01	0.96 ± 0.01	0.93 ± 0.01	2.35 ± 0.02	0.79 ± 0.19	1.85 ± 0.14
JNK2[a]	0.95 ± 0.00	0.95 ± 0.00	0.96 ± 0.00	0.91 ± 0.00	1.35 ± 0.00	1.43 ± 0.00
Mean	0.88 ± 0.02	0.95 ± 0.01	0.93 ± 0.01	1.42 ± 0.12	0.96 ± 0.13	1.49 ± 0.12

[a] Serine/threonine kinase
[b] Tyrosine kinase

Table S8. Comparison between MODBASE DFG-out models and crystal structures

	TM-Score			RMSD (Å)		
	N-lobe	C-lobe	Full domain	N-lobe	C-lobe	Full domain
ABL1[b]	0.93 ± 0.03	0.86 ± 0.06	0.89 ± 0.04	1.04 ± 0.43	1.33 ± 0.47	1.38 ± 0.51
BRAF[a]	0.93 ± 0.01	0.90 ± 0.02	0.92 ± 0.01	0.86 ± 0.16	0.64 ± 0.28	0.81 ± 0.21
KIT[b]	0.92 ± 0.06	0.95 ± 0.06	0.73 ± 0.09	1.15 ± 0.57	0.90 ± 0.42	1.88 ± 0.89
LCK[b]	0.82 ± 0.02	0.84 ± 0.02	0.83 ± 0.01	1.72 ± 0.15	1.45 ± 0.36	2.31 ± 0.37
P38α[a]	0.93 ± 0.01	0.91 ± 0.03	0.91 ± 0.02	1.06 ± 0.08	1.63 ± 0.25	1.75 ± 0.23
SRC[b]	0.87 ± 0.05	0.86 ± 0.02	0.84 ± 0.01	1.63 ± 0.50	1.47 ± 0.29	2.53 ± 0.50
EPHA3[b]	0.94 ± 0.01	0.86 ± 0.01	0.84 ± 0.01	0.58 ± 0.05	1.08 ± 0.10	2.06 ± 0.09
KDR[b]	0.92 ± 0.03	0.90 ± 0.01	0.90 ± 0.02	0.92 ± 0.43	1.07 ± 0.28	1.49 ± 0.41
CDK8[a]	0.94 ± 0.01	0.86 ± 0.01	0.89 ± 0.01	1.20 ± 0.04	0.77 ± 0.24	1.01 ± 0.13
JNK2[a]	0.94 ± 0.00	0.87 ± 0.00	0.88 ± 0.00	1.29 ± 0.00	1.37 ± 0.00	2.04 ± 0.00
Mean	0.89 ± 0.07	0.87 ± 0.05	0.87 ± 0.06	1.25 ± 0.49	1.29 ± 0.50	1.79 ± 0.61

[a] Serine/threonine kinase
[b] Tyrosine kinase

SUPPLEMENTARY REFERENCES

(1) Dar, A. C., Lopez, M. S., and Shokat, K. M. (2008) Small molecule recognition of c-Src via the Imatinib-binding conformation. *Chem. Biol. 15*, 1015-1022.

(2) Dar, A. C., Das, T. K., Shokat, K. M., and Cagan, R. L. (2012) Chemical genetic discovery of targets and anti-targets for cancer polypharmacology. *Nature 486*, 80-84.

(3) Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K., and Zarrinkar, P. P. (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol. 29*, 1046-1051.

(4) Deng, X. M., Lim, S. M., Zhang, J. M., and Gray, N. S. (2010) Broad spectrum alkynyl inhibitors of T315I Bcr-Abl. *Bioorg. Med. Chem. Lett. 20*, 4196-4200.