Supporting Information

Yin et al. 10.1073/pnas.0906146106

SI Discussion

Protein Flexibility. Our method is based on the surface features from a static structure. Therefore, surface flexibility is not addressed in this study. When the method is used to predict protein function based on surface similarity, some true positives may be missed during the search (false-negatives). How to efficiently model protein flexibility is still an open question in computational biology. One possible solution is to reduce the requirement of geometric match and incorporate physicochemical information from the surface into the fingerprints. Alternatively, starting from the unbounded structure, multiple simulations can be performed near the interface so that a set of the most preferable side-chain conformations can be explored. We will pursue these directions in future studies.

Alternative Surface Generation and Representation. The MSMS program (1) was used to generate the dot molecular surface (Connolly surface). We at first tried the University California San Francisco DMS program (http://www.cgl.ucsf.edu/Overview/software.html#dms) for dot surface generation; however, the DMS program does not distribute points uniformly on the protein surface. This resulted in the generated distribution of surface dots being irregular and not reliably representing the topology of the surface.

Additional ways exist to represent the protein surface: for example, the method based on 3D cubic grids (2). The advantage of the cubic grid is that it is easy to implement and fast to calculate. However, the surface dots will not be evenly distributed in the surface manifold. In addition, the normal vectors from each protein surface point are not well defined. For the method we adopted, we rely on the surface normal vectors to calculate the curvatures, and require relatively uniform and smooth surface features. Therefore, the dot molecular surface is a better choice.

1. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* 38:305–320.
2. Katchalskikatzir E, et al. (1992) Molecular-surface recognition—determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci USA* 89:2195–2199.

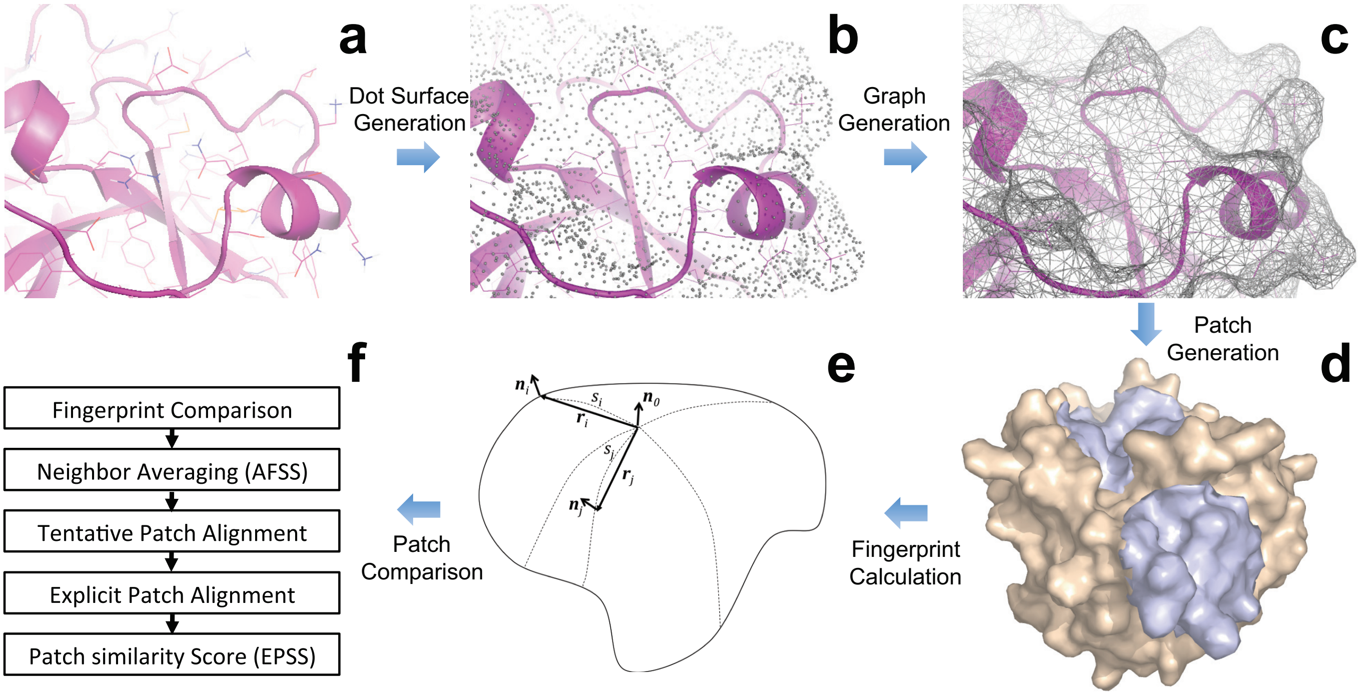

Fig. S1. Illustration of the fingerprint-based surface comparison method. From the input protein structure (*a*), a dot-surface file (*b*) is first generated using the MSMS program. (*c*) A graph representation is constructed with vertices corresponding to the surface dots, and edges connecting neighbor vertices. (Note that for illustration purposes, the number of edges has been reduced when making the figure. The actual cutoff of 2.5 Å will result in a graph with much denser connections.) (*d*) Patches are generated centering on each vertex in the graph and spanning 9 Å in geodesic distance. (*e*) We calculated the fingerprint of each patch as the geodesic distance-dependent distribution of directional curvatures measured from the center vertex. (*f*) The fingerprint-based surface patch comparison workflow (see *Methods*).

Table S1. List of the top 50 hits from screening of alpha-chymotrypsin inhibitors

PDB ID	Chain ID	EPSS score	Protein name annotation
1acb	**I**	**0**	**EGLIN C**
1cho	**I**	**47.7699**	**TURKEY OVOMUCOID THIRD DOMAIN (OMTKY3)**
1p2n	**B**	**69.7841**	**PANCREATIC TRYPSIN INHIBITOR**
2 sec	**I**	**71.2629**	**EGLIN C**
1cbw	**D**	**71.6864**	**PANCREATIC TRYPSIN INHIBITOR, BPTI**
1mtn	**D**	**71.776**	**ASIC PANCREATIC TRYPSIN INHIBITOR**
1cse	**I**	**79.6658**	**EGLIN C**
1tec	**I**	**83.7011**	**EGLIN C**
1p2n	**D**	**85.728**	**PANCREATIC TRYPSIN INHIBITOR**
1t7c	**D**	**87.2109**	**PANCREATIC TRYPSIN INHIBITOR**
2tec	**I**	**90.9058**	**EGLIN C**
1cbw	**I**	**94.0398**	**BPTI**
1okx	C	94.388	SCYPTOLIN A, PEPTIDE: 1BO-ALA-THR-THR-LEU-SUJ-CNT-VAL
1t8l	**B**	**94.4994**	**PANCREATIC TRYPSIN INHIBITOR**
1mee	**I**	**97.0394**	**EGLIN C**
1t8l	**D**	**97.9103**	**PANCREATIC TRYPSIN INHIBITOR**
2r9p	**G**	**97.9528**	**PANCREATIC TRYPSIN INHIBITOR**
1y3f	**I**	**98.0841**	**CHYMOTRYPSIN INHIBITOR 2**
1 sib	**I**	**98.8439**	**EGLIN C**
1p2q	**D**	**100.189**	**PANCREATIC TRYPSIN INHIBITOR**
1 gl1	**J**	**100.913**	**PROTEASE INHIBITOR LCMI II**
1cgj	**I**	**101.458**	**PANCREATIC SECRETORY TRYPSIN INHIBITOR (KAZAL TYPE) VARIANT 4**
1zr0	D	101.485	TISSUE FACTOR PATHWAY INHIBITOR 2
1 gl1	**I**	**103.915**	**PROTEASE INHIBITOR LCMI II**
1ca0	**I**	**104.424**	**PROTEASE INHIBITOR DOMAIN OF ALZHEIMER'S AMYLOID BETA-PROTEIN PRECURSOR**
1t8m	**B**	**104.732**	**PANCREATIC TRYPSIN INHIBITOR**
1fak	**I**	**105.915**	**PROTEIN (5L15)**
1t8m	**D**	**109.152**	**PANCREATIC TRYPSIN INHIBITOR**
1 slv	**A**	**109.709**	**ECOTIN**
2tgp	**I**	**110.175**	**TRYPSIN INHIBITOR**
1fy8	**I**	**111.249**	**PANCREATIC TRYPSIN INHIBITOR**
1yaf	A	111.568	TRANSCRIPTIONAL ACTIVATOR TENA
1ejm	**B**	**112.112**	**PANCREATIC TRYPSIN INHIBITOR**
1y4a	**I**	**113.068**	**CHYMOTRYPSIN INHIBITOR 2**
1ca0	**D**	**114.457**	**PROTEASE INHIBITOR DOMAIN OF ALZHEIMER'S AMYLOID BETA-PROTEIN PRECURSOR**
1tm3	**I**	**116.186**	**CHYMOTRYPSIN INHIBITOR 2**
1eaw	**B**	**116.702**	**PANCREATIC TRYPSIN INHIBITOR**
2buc	C	116.952	DIPEPTIDYL PEPTIDASE IV
2fi3	**I**	**116.963**	**PANCREATIC TRYPSIN INHIBITOR**
1oyv	**I**	**118.337**	**WOUND-INDUCED PROTEINASE INHIBITOR-II**
1n3t	C	119.849	GTP CYCLOHYDROLASE I
2auz	A	121.132	CATHEPSIN K
2ftm	**B**	**121.486**	**PANCREATIC TRYPSIN INHIBITOR**
1mtn	**H**	**121.744**	**BASIC PANCREATIC TRYPSIN INHIBITOR**
1lw6	**I**	**122.295**	**SUBTILISIN-CHYMOTRYPSIN INHIBITOR-2A**
1yaf	B	122.649	TRANSCRIPTIONAL ACTIVATOR TENA
1ct2	**I**	**122.865**	**OVOMUCOID INHIBITOR**
1tm7	**I**	**122.899**	**CHYMOTRYPSIN INHIBITOR 2**
1zr0	B	123.843	TISSUE FACTOR PATHWAY INHIBITOR 2
1ejm	**D**	**124.116**	**PANCREATIC TRYPSIN INHIBITOR**

The protein names are extracted from the metadata in the PDB files. The known alpha-chymotrypsin inhibitors are shown in bold.

Table S2. List of the top 20 hits from screening of uracil-DNA glycosylase inhibitors

PDB ID	Chain ID	EPSS score	Protein name annotation
1udi	**I**	**0**	**URACIL-DNA GLYCOSYLASE INHIBITOR PROTEIN**
1ugh	**I**	**32.1578**	**PROTEIN (URACIL-DNA GLYCOSYLASE INHIBITOR)**
2uug	**C**	**33.8753**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1uug	**B**	**35.9908**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
2j8x	**D**	**37.5866**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1lqm	**B**	**40.2354**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1lqg	**C**	**40.3106**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1eui	**D**	**41.3846**	**URACIL-DNA GLYCOSYLASE INHIBITOR PROTEIN**
1lqm	**F**	**43.6699**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1eui	**C**	**44.8587**	**URACIL-DNA GLYCOSYLASE INHIBITOR PROTEIN**
1uug	**D**	**45.2638**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
2j8x	**B**	**45.3427**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
2uug	**D**	**48.6658**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1lqg	**D**	**51.274i**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1lqm	**D**	**58.8174**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1lqm	**H**	**58.9295**	**URACIL-DNA GLYCOSYLASE INHIBITOR**
1yvf	A	72.5829	HCV NS5B POLYMERASE
2 h7c	C	73.6482	LIVER CARBOXYLESTERASE 1
1znj	K	91.3319	INSULIN
2e7l	A	92.1639	T CELL RECEPTOR ALPHA CHAIN

The protein names are extracted from the metadata in the PDB files. The known uracil-DNA glycosylase inhibitors are shown in bold.

Table S3. List of top 50 hits from screening of estrogen receptor

PDB ID	Chain ID	EPSS Score	Protein name annotation
1qkn	**A**	**0**	**ESTROGEN RECEPTOR BETA**
1hj1	**A**	**39.7703**	**ESTROGEN RECEPTOR BETA**
2ayr	**A**	**40.4564**	**ESTROGEN RECEPTOR**
1xp1	**A**	**42.8517**	**ESTROGEN RECEPTOR**
1yim	**A**	**43.1529**	**ESTROGEN RECEPTOR**
2bj4	**B**	**43.7577**	**ESTROGEN RECEPTOR ALPHA**
1xp9	**A**	**44.2426**	**ESTROGEN RECEPTOR**
2jf9	**A**	**44.9503**	**ESTROGEN RECEPTOR**
1xp6	**A**	**45.4521**	**ESTROGEN RECEPTOR**
1xpc	**A**	**45.7676**	**ESTROGEN RECEPTOR**
2ouz	**A**	**46.361**	**ESTROGEN RECEPTOR**
1gz7	A	47.0982	LIPASE 2
1o0s	B	48.393	NAD-DEPENDENT MALIC ENZYME
2jfa	**A**	**50.4815**	**ESTROGEN RECEPTOR**
1a2z	B	50.6728	PYRROLIDONE CARBOXYL PEPTIDASE
2q70	**A**	**50.7383**	**ESTROGEN RECEPTOR**
1n23	B	51.1064	(+)-BORNYL DIPHOSPHATE SYNTHASE
2gpu	**A**	**52.059**	**ESTROGEN-RELATED RECEPTOR GAMMA**
1a52	**B**	**55.9095**	**ESTROGEN RECEPTOR**
1lpm	A	56.2002	LIPASE
1r5k	**B**	**56.6413**	**ESTROGEN RECEPTOR**
1gz7	D	57.0936	LIPASE 2
1e1f	A	57.7041	BETA-GLUCOSIDASE
1llq	B	57.9091	NAD-DEPENDENT MALIC ENZYME
1bpq	A	58.3395	PHOSPHOLIPASE A2
2cjf	C	58.9727	3-DEHYDROQUINATE DEHYDRATASE
2fsz	**B**	**59.053**	**ESTROGEN RECEPTOR BETA**
1e1e	B	60.0215	BETA-GLUCOSIDASE
2jfa	**B**	**60.0931**	**ESTROGEN RECEPTOR**
1ker	A	60.7205	DTDP-D-GLUCOSE 4,6-DEHYDRATASE
1err	**B**	**60.8233**	**ESTROGEN RECEPTOR**
1r5k	**C**	**61.008**	**ESTROGEN RECEPTOR**
2qgv	H	61.3767	HYDROGENASE-1 OPERON PROTEIN HYAE
2i0j	**B**	**61.8803**	**ESTROGEN RECEPTOR ALPHA**
1yb4	A	62.6156	TARTRONIC SEMIALDEHYDE REDUCTASE
2qe4	**B**	**62.8829**	**ESTROGEN RECEPTOR**
2dhz	A	63.5922	RAP GUANINE NUCLEOTIDE EXCHANGE FACTOR (GEF)- LIKE 1
2pgj	B	64.7522	ADP-RIBOSYL CYCLASE 1
1gkq	D	65.4463	HYDANTOINASE
1xot	A	65.7045	CAMP-SPECIFIC 3′,5′-CYCLIC PHOSPHODIESTERASE 4B
1tb7	B	65.8592	CAMP-SPECIFIC 3′,5′-CYCLIC PHOSPHODIESTERASE 4D
1h27	C	65.9094	CELL DIVISION PROTEIN KINASE 2
2gkl	A	65.9506	BETA-LACTAMASE
1wmk	A	66.0729	DEATH-ASSOCIATED PROTEIN KINASE 2
2hqa	A	67.211	DNA POLYMERASE III ALPHA SUBUNIT
1yin	**A**	**67.4511**	**ESTROGEN RECEPTOR**
1gow	B	67.836	BETA-GLYCOSIDASE
2jf9	**C**	**67.8658**	**ESTROGEN RECEPTOR**
2pmb	A	68.0196	UNCHARACTERIZED PROTEIN
1a52	**A**	**68.2207**	**ESTROGEN RECEPTOR**

Protein names are extracted from the metadata in the PDB files. Known estrogen receptors are shown in bold. The query structure (PDB ID: 1qkn) is shown as a trivial hit from the screening.

Table S4. List of top 50 hits from screening of cyclin-dependent kinase 2

PDB ID	Chain ID	EPSS Score	Protein Name Annotation
1di8	**A**	**0**	**CYCLIN-DEPENDENT KINASE 2**
1oiy	**C**	**38.161**	**CELL DIVISION PROTEIN KINASE 2**
2iw8	**C**	**40.3031**	**CELL DIVISION PROTEIN KINASE 2**
1aq1	**A**	**45.0125**	**CYCLIN-DEPENDENT PROTEIN KINASE 2**
1fvt	**A**	**45.8081**	**CELL DIVISION PROTEIN KINASE 2**
1oi9	**C**	**47.3152**	**CELL DIVISION PROTEIN KINASE 2**
1ke9	**A**	**47.4172**	**CELL DIVISION PROTEIN KINASE 2**
2uue	**C**	**49.434**	**CELL DIVISION PROTEIN KINASE 2**
1ogu	**A**	**50.5897**	**CELL DIVISION PROTEIN KINASE 2**
2i40	**C**	**50.876**	**CELL DIVISION PROTEIN KINASE 2**
1ckp	**A**	**51.0618**	**PROTEIN (CYCLIN-DEPENDENT PROTEIN KINASE 2)**
1 h1p	**A**	**52.4423**	**CELL DIVISION PROTEIN KINASE 2**
1e9 h	**C**	**52.4473**	**CELL DIVISION PROTEIN KINASE 2**
1 h1q	**C**	**53.0548**	**CELL DIVISION PROTEIN KINASE 2**
1oiu	**C**	**53.4459**	**CELL DIVISION PROTEIN KINASE 2**
1v1k	**A**	**54.7113**	**CELL DIVISION PROTEIN KINASE 2**
1ke8	**A**	**54.8103**	**CELL DIVISION PROTEIN KINASE 2**
1ogu	**C**	**54.9141**	**CELL DIVISION PROTEIN KINASE 2**
2uzl	**C**	**54.9601**	**CELL DIVISION PROTEIN KINASE 2**
1oi9	**A**	**55.2998**	**CELL DIVISION PROTEIN KINASE 2**
2c5n	**A**	**55.776**	**CELL DIVISION PROTEIN KINASE 2**
2iw6	**C**	**56.7318**	**CELL DIVISION PROTEIN KINASE 2**
2uzl	**A**	**57.4596**	**CELL DIVISION PROTEIN KINASE 2**
2iw6	**A**	**57.5462**	**CELL DIVISION PROTEIN KINASE 2**
2c5o	**A**	**58.147**	**CELL DIVISION PROTEIN KINASE 2**
1 h26	**C**	**58.9842**	**CELL DIVISION PROTEIN KINASE 2**
2iw9	**C**	**59.5102**	**CELL DIVISION PROTEIN KINASE 2**
2c4 g	**C**	**60.2306**	**CELL DIVISION PROTEIN KINASE 2**
2iw8	**A**	**61.1751**	**CELL DIVISION PROTEIN KINASE 2**
1vyw	**A**	**62.3481**	**CELL DIVISION PROTEIN KINASE 2**
1 h1r	**C**	**62.4581**	**CELL DIVISION PROTEIN KINASE 2**
1ol2	**C**	**62.9078**	**CELL DIVISION PROTEIN KINASE 2**
1e9 h	**A**	**63.2369**	**CELL DIVISION PROTEIN KINASE 2**
2iw9	**A**	**63.2699**	**CELL DIVISION PROTEIN KINASE 2**
2duv	**A**	**63.8217**	**CELL DIVISION PROTEIN KINASE 2**
1oiu	**A**	**64.1929**	**CELL DIVISION PROTEIN KINASE 2**
2uzd	**A**	**65.0139**	**CELL DIVISION PROTEIN KINASE 2**
2b52	**A**	**65.4456**	**CELL DIVISION PROTEIN KINASE 2**
1ykr	**A**	**65.5259**	**CELL DIVISION PROTEIN KINASE 2**
2c5n	**C**	**65.5853**	**CELL DIVISION PROTEIN KINASE 2**
1oiy	**A**	**66.3882**	**CELL DIVISION PROTEIN KINASE 2**
1oir	**A**	**67.0979**	**CELL DIVISION PROTEIN KINASE 2**
2uzb	**A**	**68.0866**	**CELL DIVISION PROTEIN KINASE 2**
1 sm2	B	68.2988	TYROSINE-PROTEIN KINASE ITK/TSK
2f2c	**B**	**68.741**	**CELL DIVISION PROTEIN KINASE 6**
2uze	**C**	**69.5631**	**CELL DIVISION PROTEIN KINASE 2**
2cch	**A**	**69.8073**	**CELL DIVISION PROTEIN KINASE 2**
1 h07	**A**	**70.6655**	**CELL DIVISION PROTEIN KINASE 2**
2uzn	**A**	**71.8319**	**CELL DIVISION PROTEIN KINASE 2**
2oib	A	72.0989	INTERLEUKIN-1 RECEPTOR-ASSOCIATED KINASE 4

Protein names are extracted from the metadata in the PDB files. Known cyclin-dependent kinase 2 proteins are shown in bold. The query structure (PDBID: 1di8) is also shown as the top trivial hit from the screening.

Table S5. Ranking and scoring of the 243 selected chymotrypsin inhibitors during screening of the PDB

PDB ID	Rank	EPSS
1acb-I	1	0
1cho-I	2	47.7699
1p2n-B	3	69.7841
2 sec-I	4	71.2629
1cbw-D	5	71.6864
1mtn-D	6	71.776
1cse-I	7	79.6658
1tec-I	8	83.7011
1p2n-D	9	85.728
1t7c-D	10	87.2109
2tec-I	11	90.9058
1cbw-I	12	94.0398
1t8l-B	14	94.4994
1mee-I	15	97.0394
1t8l-D	16	97.9103
2r9p-G	17	97.9528
1y3f-I	18	98.0841
1 sib-I	19	98.8439
1p2q-D	20	100.189
1 gl1-J	21	100.913
1 gl1-I	24	103.915
1ca0-I	25	104.424
1t8m-B	26	104.732
1fak-I	27	105.915
1t8m-D	28	109.152
1 slv-A	29	109.709
2tgp-I	30	110.175
1fy8-I	31	111.249
1ejm-B	33	112.112
1y4a-I	34	113.068
1ca0-D	35	114.457
1tm3-I	36	116.186
1eaw-B	37	116.702
2ftm-B	43	121.486
1mtn-H	44	121.744
1lw6-I	45	122.295
1ct2-I	47	122.865
1tm7-I	48	122.899
1ejm-D	50	124.116
1bpt-A	51	124.214
1t7c-B	53	124.742
1 slw-A	55	125.279
2kai-I	57	125.591
1to1-I	59	126.082
1to2-I	61	126.78
1 gl1-K	62	127.119
1bth-Q	65	129.445
1f5r-I	66	129.77
1t8n-D	67	130.258
1t8n-B	68	130.862
1p2o-B	71	131.686
1 slx-A	73	132.468
1p2o-D	76	132.982
2tpi-I	77	133.245
1r0r-I	85	137.05
1tmg-I	88	137.486
1brc-I	91	138.496
1y4d-I	92	138.595
1bzx-I	94	138.947
1p2q-B	96	139.381
2 sgd-I	100	140.387
1cso-I	103	141.327
1 slu-A	104	141.327
2 sni-I	105	141.494
1ppf-I	110	143.227
1yu6-D	113	143.884
1xx9-C	117	144.66
1co7-I	119	145.11
1t8o-D	121	145.327
1azz-C	124	145.788
1xxf-C	126	147.515
1brb-I	127	147.743
1ejm-F	129	148.084
1t8o-B	131	148.634
1 sbn-I	135	149.961
1y1k-I	146	153.127
1y3b-I	147	153.468
1taw-B	153	154.844
1xxf-D	157	155.792
1p2j-I	163	157.32
1bth-P	169	158.014
1tm1-I	176	159.102
1tx6-I	186	160.495
1tpa-I	190	160.757
1 sgy-I	191	161.231
1 hja-I	218	165.227
1 sge-I	223	165.625
1z7k-B	228	166.2
2ptc-I	229	166.297
2ftl-I	231	166.728
1tm5-I	237	167.087
1ds2-I	240	167.378
1p0 s-E	245	167.826
1 sgr-I	248	168.473
1ezs-B	250	168.644
2 sge-I	266	170.008
1ezs-A	287	172.51
2 gkr-I	294	173.226
1 sgd-I	296	173.595
1tab-I	297	173.649
1 gl0-I	313	175.547
1y48-I	341	178.355
1ct4-I	383	181.399
1 sgn-I	386	181.571
2 sgq-I	461	186.682
1f7z-I	462	186.805
2 sgf-I	485	188.245
1id5-I	509	189.605
1y3c-I	543	191.991
1y34-I	544	192.01
1omu-A	581	193.742
1y3d-I	596	194.625
1y33-I	615	196.471
1yu6-C	679	199.556
2ijo-I	701	200.427
1p2m-B	862	206.536
1nag-A	925	208.778
1ykt-B	1,167	216.993
1tm4-I	1,175	217.116
1ezu-A	1,394	222.362
1iy5-A	1,640	228.146
1ds3-I	1,861	232.617
1 h34-A	2,114	236.862
1p2m-D	2,183	237.913
1n8o-E	2,318	239.942
1 g6x-A	2,481	242.457
1ce3-A	2,509	242.833
1ypb-I	2,701	245.716

PDB ID	Rank	EPSS
1ct0-I	2,806	247.095
1aap-A	2,964	249.067
2 sgp-I	3,439	254.406
1ypa-I	3,633	256.52
1ciq-A	3,693	257.069
1ovo-D	4,255	262.148
1ypc-I	4,809	266.628
1eaw-D	4,839	266.92
1 sgp-I	5,138	269.356
1bpi-A	5,169	269.628
1ovo-C	6,281	277.476
1xx9-D	6,491	278.858
1uub-A	6,920	281.466
1cq4-A	8,295	289.119
1azz-D	8,485	290.153
1k6u-A	9,529	295.358
2iln-I	9,909	297.009
2nu3-I	10,631	300.095
1eai-C	11,611	304.113
1 sgq-I	13,676	311.749
1iy6-A	14,442	314.52
1b0c-E	17,265	323.481
1bhc-A	17,560	324.462
1b0c-B	19,938	331.195
1bhc-H	20,741	333.396
1bz5-E	21,168	334.397
1bhc-G	21,854	336.156
1bhc-C	21,931	336.374
1bz5-C	22,362	337.477
1bhc-B	22,703	338.343
1jv9-A	23,407	340.067
1b0c-A	23,923	341.324
2 hex-A	24,298	342.223
1fan-A	24,336	342.315
1bbi-A	25,297	344.66
1tur-A	25,430	344.933
1bhc-D	25,937	346.167
1p2i-I	26,142	346.638
1bhc-E	26,883	348.343
1wo9-A	26,950	348.5
1bi6-H	27,180	349.016
1df9-C	27,267	349.213
1omt-A	27,543	349.859
1bhc-J	27,933	350.771
1coa-I	28,156	351.327
2ci2-I	28,309	351.66
1c2a-A	29,016	353.229
1ecy-A	30,366	356.18
2ftm-A	30,613	356.687
1ovo-A	32,038	359.659
1bz5-B	32,717	361.001
1mvz-A	32,882	361.365
1bhc-I	34,710	365.16
1bti-A	35,000	365.774
1bhc-F	37,015	369.829
1fn0-A	37,047	369.868
1qh2-B	38,489	372.79
1ifg-A	39,667	374.957
2 hex-B	39,891	375.369
2bbi-A	42,172	379.75
1zjd-B	43,153	381.665
1eai-D	44,552	384.275
1bz5-A	45,509	386.073
1ccv-A	45,660	386.344
1m8b-A	46,122	387.238
1jxc-A	46,642	388.182
1pit-A	49,019	392.499
2 hex-C	49,304	392.935
1pbi-B	51,046	396.051
2ovo-A	51,236	396.381
1ecz-A	51,460	396.823
1aal-A	51,676	397.196
1egl-A	53,792	401.041
1b0c-D	54,968	403.215
1qlq-A	56,912	406.779
2wbc-A	58,238	409.197
1aal-B	58,239	409.198
1ecz-B	58,717	410.073
2 hex-D	62,502	417.083
1bi6-L	62,612	417.295
1ovo-B	65,927	423.58
2bi6-H	65,971	423.672
1 hx2-A	66,246	424.17
1fi8-E	71,220	434.226
1k9b-A	74,305	440.566
1fyb-A	75,490	443.183
1kio-A	75,806	443.858
1xxd-C	76,048	444.341
1m8c-A	76,301	444.915
1eyl-A	76,652	445.619
2fj8-A	79,477	452.163
1pi2-A	81,823	457.96
1tih-A	83,799	463.158
1fmz-A	83,974	463.599
1d0d-B	84,576	465.149
1ezu-B	85,829	468.656
1b0c-C	86,655	470.977
1pbi-A	87,142	472.482
1uua-A	88,395	476.216
2 hex-E	89,222	478.883
1aap-B	90,714	483.926
1xxd-D	91,135	485.405
1jv8-A	92,282	489.827
1fi8-C	93,004	492.688
1cir-A	93,403	494.445
2bi6-L	93,895	496.461
1kj0-A	95,633	504.173
1egp-A	96,013	505.927
1p2k-I	97,009	510.801
1d6r-I	97,424	513.188
1tus-A	97,707	514.732
1bz5-D	97,973	516.344
1wbc-A	99,985	529.412
1kgm-A	105,325	605.445
1pmc-A	105,615	615.192

Table S6. Ranking and scoring of the 26 selected uracil-DNA glycosylase inhibitors during screening of the PDB

PDB ID	RANK	EPSS
1udi-I	1	0
1ugh-I	2	32.1578
2uug-C	3	33.8753
1uug-B	4	35.9908
2j8x-D	5	37.5866
1lqm-B	6	40.2354
1lqg-C	7	40.3106
1eui-D	8	41.3846
1lqm-F	9	43.6699
1eui-C	10	44.8587
1uug-D	11	45.2638
2j8x-B	12	45.3427
2uug-D	13	48.6658
1lqg-D	14	51.274
1lqm-D	15	58.8174
1lqm-H	16	58.9295
1ugi-A	9,432	214.758
2ugi-B	16,532	236.823
1ugi-D	17,085	238.319
1ugi-C	17,877	240.364
1ugi-F	19,343	244.09
1ugi-G	46,704	302.665
1ugi-E	74,234	361.319
1ugi-H	78,997	374.152
2ugi-A	85,958	395.591
1ugi-B	89,364	408.298