

Supporting Information for

Conformational Sampling of Drug-like Molecules with MOE and Catalyst: Implications
for Pharmacophore Modelling and Virtual Screening

Table S1. List of proteins-ligand complexes for the three compound sets (Bostrom, modified-Perola, Vernalis) used in the present study. Entries within a dataset are sorted on the PDB ID. There are 256 unique compounds, but 262 entries to the table because there are 6 compounds in common between the Bostrom and modified-Perola sets.

PDB ID	Protein	X-ray Resolution (Å)	Set
1a28	PROGESTERONE RECEPTOR	1.80	Bostrom
1bju	BETA-TRYPSIN	1.80	Bostrom
1bzs	NEUTROPHIL COLLAGENASE (MMP8)	1.70	Bostrom
1c83	TYROSINE PHOSPHATASE 1B	1.80	Bostrom
1caq	STROMELYSIN	1.80	Bostrom
1cbs	RETINOIC-ACID-BINDING PROTEIN	1.80	Bostrom
1cbx	CARBOXYPEPTIDASE A	2.00	Bostrom
1d3g	DIHYDROOROTATE DEHYDROGENASE	1.60	Bostrom
1dam	DETHIOBIOTIN SYNTHETASE	1.80	Bostrom
1dyr	DIHYDROFOLATE REDUCTASE	1.86	Bostrom
1ecv	TYROSINE PHOSPHATASE 1B	1.95	Bostrom
1ejn	UROKINASE	1.80	Bostrom
1f0u	TRYPSIN	1.90	Bostrom
1fcz	RETINOIC ACID NUCLEAR RECEPTOR	1.38	Bostrom
1fkg	FK506 BINDING PROTEIN	2.00	Bostrom
1frb	FR-1 PROTEIN	1.70	Bostrom
1ftm	GLUTAMATE RECEPTOR-2	1.70	Bostrom
1gr2	GLUTAMATE RECEPTOR-2	1.90	Bostrom
1ian	P38 MAP KINASE	2.00	Bostrom
1mtv	TRYPSIN	1.90	Bostrom
1mtw	TRYPSIN	1.90	Bostrom
1phg	CYTOCHROME P450-CAM	1.60	Bostrom
1ppc	TRYPSIN	1.80	Bostrom
1pph	TRYPSIN	1.90	Bostrom
1qft	HISTAMINE BINDING PROTEIN	1.25	Bostrom
1tng	TRYPSIN	1.80	Bostrom
1tnh	TRYPSIN	1.80	Bostrom
2izg	STREPTAVIDIN	1.36	Bostrom
3bto	LIVER ALCOHOL DEHYDROGENASE	1.66	Bostrom
5std	SCYTALONE DEHYDRATASE	1.95	Bostrom
6std	SCYTALONE DEHYDRATASE	1.80	Bostrom
7std	SCYTALONE DEHYDRATASE	1.80	Bostrom
13gs	GLUTATHIONE S-TRANSFERASE	1.90	Perola
1a42	CARBONIC ANHYDRASE II	2.25	Perola
1a4k	ANTIBODY FAB	2.40	Perola
1a8t	METALLO-BETA-LACTAMASE	2.55	Perola
1afq	BOVINE GAMMA-CHYMOTRYPSIN	1.80	Perola
1aoe	DIHYDROFOLATE REDUCTASE	1.60	Perola
1atl	ATROLYSIN C	1.80	Perola
1azm	LYASE(OXO-ACID)	2.00	Perola
1bnw	CARBONIC ANHYDRASE	2.25	Perola
1bqo	STROMELYSIN-1	2.30	Perola
1br6	RICIN	2.30	Perola
1cet	L-LACTATE DEHYDROGENASE	2.05	Perola

1cim	LYASE(OXO-ACID)	2.10	Perola
1d3p	ALPHA-THROMBIN	2.10	Perola
1d4p	ALPHA-THROMBIN	2.07	Perola
1d6v	CATALYTIC ANTIBODY AZ-28	2.00	Perola
1dib	METHYLENETETRAHYDROFOLATE	2.70	Perola
1dlr	OXIDO-REDUCTASE	2.30	Perola
1efy	POLY (ADP-RIBOSE) POLYMERASE	2.20	Perola
1ela	HYDROLASE(SERINE PROTEINASE)	1.80	Perola
1etr	HYDROLASE(SERINE PROTEINASE)	2.20	Perola
1ett	HYDROLASE(SERINE PROTEINASE)	2.50	Perola
1eve	ACETYLCHOLINESTERASE	2.50	Perola
1exa	RETINOIC ACID RECEPTOR GAMMA-2	1.59	Perola
1ezq	COAGULATION FACTOR XA	2.20	Perola
1f0r	COAGULATION FACTOR XA	2.10	Perola
1f0t	TRYPSIN	1.80	Perola
1f4e	THYMIDYLATE SYNTHASE	1.90	Perola
1f4f	THYMIDYLATE SYNTHASE	2.00	Perola
1f4g	THYMIDYLATE SYNTHASE	1.75	Perola
1fcx	RETINOIC ACID RECEPTOR GAMMA-1	1.47	Perola
1fcz	RETINOIC ACID RECEPTOR GAMMA-1	1.38	Perola
1fjs	COAGULATION FACTOR XA	1.92	Perola
1fkg	CIS-TRANS ISOMERASE	2.00	Perola
1fm6	RETINOIC ACID RECEPTOR RXR-ALPHA	2.10	Perola
1fm9	RETINOIC ACID RECEPTOR RXR-ALPHA	2.10	Perola
1frb	FR-1 PROTEIN	1.70	Perola
1g4o	CARBONIC ANHYDRASE II	1.96	Perola
1gwx	PPAR-DELTA	2.50	Perola
1h1p	CYCLIN-DEPENDENT KINASE II	2.10	Perola
1h1s	CYCLIN-DEPENDENT KINASE II	2.00	Perola
1h9u	RETINOID X RECEPTOR	2.70	Perola
1hdq	CARBOXYPEPTIDASEA	2.30	Perola
1hfc	METALLOPROTEASE	1.56	Perola
1hpv	HYDROLASE (ACID PROTEINASE)	1.90	Perola
1htf	HYDROLASE(ACID PROTEINASE)	2.20	Perola
1i7z	ANTIBODY GNC92H2	2.30	Perola
1i8z	CARBONIC ANHYDRASE II	1.93	Perola
1if7	CARBONIC ANHYDRASE II	1.98	Perola
1iy7	CARBOXYPEPTIDASE A	2.00	Perola
1jsv	CYCLIN-DEPENDENT KINASE II	1.96	Perola
1k1j	TRYPSIN	2.20	Perola
1k22	PROTHROMBIN	1.93	Perola
1k7e	TRYPTOPHAN SYNTHASE ALPHA CHAIN	2.30	Perola
1k7f	TRYPTOPHAN SYNTHASE ALPHA CHAIN	1.90	Perola
1kv1	P38 MAP KINASE	2.50	Perola
1kv2	P38 MAP KINASE	2.80	Perola
1l2s	BETA-LACTAMASE	1.94	Perola
1l8g	PROTEIN-TYROSINE PHOSPHATASE	2.50	Perola
1lqd	BLOOD COAGULATION FACTOR XA	2.70	Perola
1m48	INTERLEUKIN-2	1.95	Perola
1mmb	NEUTROPHIL COLLAGENASE	2.10	Perola
1mnc	HYDROLASE (METALLOPROTEASE)	2.10	Perola
1mq5	COAGULATION FACTOR X	2.10	Perola

1mq6	COAGULATION FACTOR X	2.10	Perola
1nhu	HCV RNA POLYMERASE	2.00	Perola
1nhv	HCV RNA POLYMERASE	2.90	Perola
1o86	ANGIOTENSIN CONVERTING ENZYME	2.00	Perola
1ohr	ASPARTYLPROTEASE	2.10	Perola
1ppc	HYDROLASE(SERINE PROTEINASE)	1.80	Perola
1pph	HYDROLASE(SERINE PROTEINASE)	1.90	Perola
1qbu	HIV-1 PROTEASE	1.80	Perola
1qhi	THYMIDINE KINASE	1.90	Perola
1ql9	TRYPSIN	2.30	Perola
1qpe	LCK KINASE	2.00	Perola
1r09	RHINOVIRUS COAT PROTEIN	2.90	Perola
1syn	THYMIDYLATE SYNTHASE	2.00	Perola
1thl	THERMOLYSIN	1.70	Perola
1uvs	THROMBIN	2.80	Perola
1uvt	THROMBIN	2.50	Perola
1ydr	C-AMP-DEPENDENT PROTEIN KINASE	2.20	Perola
1yds	C-AMP-DEPENDENT PROTEIN KINASE	2.20	Perola
1ydt	C-AMP-DEPENDENT PROTEIN KINASE	2.30	Perola
2cgr	IMMUNOGLOBULIN	2.20	Perola
2csn	CASEIN KINASE-1	2.50	Perola
2pcp	IMMUNOGLOBULIN	2.20	Perola
2qwi	NEURAMINIDASE	2.00	Perola
3cpa	HYDROLASE (C-TERMINAL PEPTIDASE)	2.00	Perola
3erk	EXTRACELLULAR REGULATED KINASE 2	2.10	Perola
3ert	ESTROGEN RECEPTOR ALPHA	1.90	Perola
3std	SCYTALONE DEHYDRATASE	1.65	Perola
3tmn	HYDROLASE (METALLOPROTEINASE)	1.70	Perola
4dfr	OXIDO-REDUCTASE	1.70	Perola
4std	SCYTALONE DEHYDRATASE	2.15	Perola
5std	SCYTALONE DEHYDRATASE	1.95	Perola
5tln	HYDROLASE (METALLOPROTEINASE)	2.30	Perola
7dfr	OXIDO-REDUCTASE	2.50	Perola
7est	HYDROLASE (SERINE PROTEINASE)	1.80	Perola
830c	COLLAGENASE-3 (MMP-13)	1.60	Perola
966c	COLLAGENASE-1 (MMP-1)	1.90	Perola
1bl6	P38 MAP KINASE	2.50	Vernalis
1bmK	P38 MAP KINASE	2.40	Vernalis
1bx6	C-AMP DEPENDENT KINASE	2.10	Vernalis
1clc	HIV-1 REVERSE TRANSCRIPTASE	2.50	Vernalis
1di8	CYCLIN-DEPENDENT KINASE II	2.20	Vernalis
1elv	CYCLIN-DEPENDENT KINASE II	1.95	Vernalis
1e9h	CYCLIN-DEPENDENT KINASE II-CYCLIN	2.50	Vernalis
1ep4	HIV-1 REVERSE TRANSCRIPTASE	2.50	Vernalis
1fk9	HIV-1 REVERSE TRANSCRIPTASE	2.50	Vernalis
1fpu	ABL KINASE	2.40	Vernalis
1fvt	CYCLIN-DEPENDENT KINASE II	2.20	Vernalis
1gii	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1gij	CYCLIN-DEPENDENT KINASE II	2.20	Vernalis
1gz8	CYCLIN-DEPENDENT KINASE II	1.30	Vernalis
1h07	CYCLIN-DEPENDENT KINASE II	1.85	Vernalis
1h0v	CYCLIN-DEPENDENT KINASE II	1.90	Vernalis

1h0w	CYCLIN-DEPENDENT KINASE II	2.10	Vernalis
1h1r	CYCLIN-DEPENDENT KINASE II-CYCLIN	2.00	Vernalis
1iep	ABL KINASE	2.10	Vernalis
1jla	HIV-1 REVERSE TRANSCRIPTASE	2.50	Vernalis
1ke5	CYCLIN-DEPENDENT KINASE II	2.20	Vernalis
1ke6	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1ke7	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1ke8	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1ke9	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1m7q	P38 MAP KINASE	2.40	Vernalis
1o9u	GLYCOGEN SYNTHASE KINASE 3 BETA	2.40	Vernalis
1oi9	CYCLIN-DEPENDENT KINASE II	2.10	Vernalis
1oiq	CYCLIN-DEPENDENT KINASE II	2.31	Vernalis
1oir	CYCLIN-DEPENDENT KINASE II	1.91	Vernalis
1oit	CYCLIN-DEPENDENT KINASE II	1.60	Vernalis
1opk	ABL KINASE	1.80	Vernalis
1ouk	P38 MAP KINASE	2.50	Vernalis
1ouy	P38 MAP KINASE	2.50	Vernalis
1ove	P38 MAP KINASE	2.10	Vernalis
1oz1	P38 MAP KINASE	2.10	Vernalis
1p2a	CYCLIN-DEPENDENT KINASE II	2.50	Vernalis
1pmq	MITOGEN-ACTIVATED PROTEIN KINASE 10	2.20	Vernalis
1pxi	CYCLIN-DEPENDENT KINASE II	1.95	Vernalis
1pxj	CYCLIN-DEPENDENT KINASE II	2.30	Vernalis
1pxl	CYCLIN-DEPENDENT KINASE II	2.50	Vernalis
1pxn	CYCLIN-DEPENDENT KINASE II	2.50	Vernalis
1pxp	CYCLIN-DEPENDENT KINASE II	2.30	Vernalis
1py5	TGF-BETA RECEPTOR TYPE I	2.30	Vernalis
1pye	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1q3w	GLYCOGEN SYNTHASE KINASE 3 BETA	2.30	Vernalis
1q5k	GLYCOGEN SYNTHASE KINASE 3 BETA	1.94	Vernalis
1q8t	C-AMP DEPENDENT KINASE	2.00	Vernalis
1q8w	C-AMP DEPENDENT KINASE	2.20	Vernalis
1qkn	ESTROGEN RECEPTOR BETA	2.25	Vernalis
1qs4	HIV-1 INTEGRASE	2.10	Vernalis
1r0e	GLYCOGEN SYNTHASE KINASE 3 BETA	2.25	Vernalis
1r78	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1rek	C-AMP DEPENDENT KINASE	2.30	Vernalis
1rw8	TGF-BETA RECEPTOR TYPE I	2.40	Vernalis
1s9p	ESTROGEN RECEPTOR GAMMA	2.13	Vernalis
1s1t	HIV-1 REVERSE TRANSCRIPTASE	2.40	Vernalis
1u3q	ESTROGEN RECEPTOR BETA	2.40	Vernalis
1ung	CYCLIN-DEPENDENT KINASE V	2.30	Vernalis
1unl	CYCLIN-DEPENDENT KINASE V	2.20	Vernalis
1uom	ESTROGEN RECEPTOR	2.28	Vernalis
1urw	CYCLIN-DEPENDENT KINASE II	1.60	Vernalis
1uy6	HEAT SHOCK PROTEIN 90	1.90	Vernalis
1uy7	HEAT SHOCK PROTEIN 90	1.90	Vernalis
1uyc	HEAT SHOCK PROTEIN 90	2.00	Vernalis
1uyd	HEAT SHOCK PROTEIN 90	2.20	Vernalis
1uyf	HEAT SHOCK PROTEIN 90	2.00	Vernalis
1uyg	HEAT SHOCK PROTEIN 90	2.00	Vernalis

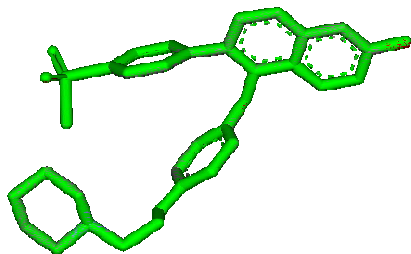
1uyh	HEAT SHOCK PROTEIN 90	2.20	Vernalis
1uyi	HEAT SHOCK PROTEIN 90	2.20	Vernalis
1uyk	HEAT SHOCK PROTEIN 90	2.20	Vernalis
1v0p	P. FALCIPARUM PROTEIN KINASE 5	2.00	Vernalis
1v1k	CYCLIN-DEPENDENT KINASE II	2.31	Vernalis
1vjy	TGF-BETA RECEPTOR TYPE I	2.00	Vernalis
1vyw	CYCLIN-DEPENDENT KINASE II	2.30	Vernalis
1vyz	CYCLIN-DEPENDENT KINASE II	2.21	Vernalis
1w0x	CYCLIN-DEPENDENT KINASE II	2.20	Vernalis
1w83	P38 MAP KINASE	2.50	Vernalis
1w84	P38 MAP KINASE	2.20	Vernalis
1wbn	P38 MAP KINASE	2.40	Vernalis
1wbs	P38 MAP KINASE	1.80	Vernalis
1wbw	P38 MAP KINASE	2.41	Vernalis
1wzy	MITOGEN-ACTIVATED PROTEIN KINASE 1	2.50	Vernalis
1x76	ESTROGEN RECEPTOR BETA	2.20	Vernalis
1x78	ESTROGEN RECEPTOR BETA	2.30	Vernalis
1x8b	WEE1-LIKE PROTEIN KINASE	1.81	Vernalis
1xp1	ESTROGEN RECEPTOR ALPHA	1.80	Vernalis
1xqc	ESTROGEN RECEPTOR ALPHA	2.05	Vernalis
1xws	PIM1 KINASE DOMAIN	1.80	Vernalis
1y8y	CYCLIN-DEPENDENT KINASE II	2.00	Vernalis
1y91	CYCLIN-DEPENDENT KINASE II	2.15	Vernalis
1yc3	HEAT SHOCK PROTEIN 90	2.12	Vernalis
1yi3	PIM1 KINASE DOMAIN	2.50	Vernalis
1yim	ESTROGEN RECEPTOR ALPHA	1.90	Vernalis
1yin	ESTROGEN RECEPTOR ALPHA	2.20	Vernalis
1ykr	CYCLIN-DEPENDENT KINASE II	1.80	Vernalis
1yqj	P38 MAP KINASE	2.00	Vernalis
1yw2	P38 MAP KINASE	2.01	Vernalis
1ywr	P38 MAP KINASE	1.95	Vernalis
1yxx	PIM1 KINASE DOMAIN	2.00	Vernalis
1z5m	3-PHOSPHOINOSITIDE-DEPENDENT PROTEIN KINASE 1	2.17	Vernalis
1zyj	P38 MAP KINASE	2.00	Vernalis
1zz2	P38 MAP KINASE	2.00	Vernalis
1zzl	P38 MAP KINASE	2.00	Vernalis
2a0c	CYCLIN-DEPENDENT KINASE II	1.95	Vernalis
2ayr	ESTROGEN RECEPTOR	1.90	Vernalis
2b52	CYCLIN-DEPENDENT KINASE II	1.88	Vernalis
2bak	P38 MAP KINASE	2.20	Vernalis
2bal	P38 MAP KINASE	2.10	Vernalis
2be2	HIV-1 REVERSE TRANSCRIPTASE	2.43	Vernalis
2bfy	AURORA KINASE B	1.80	Vernalis
2br1	CHECKPOINT KINASE 1	2.00	Vernalis
2brb	CHECKPOINT KINASE 1	2.10	Vernalis
2brc	HEAT SHOCK PROTEIN 90	1.60	Vernalis
2bre	HEAT SHOCK PROTEIN 90	2.00	Vernalis
2brg	CHECKPOINT KINASE 1	2.10	Vernalis
2brh	CHECKPOINT KINASE 1	2.10	Vernalis
2brm	CHECKPOINT KINASE 1	2.20	Vernalis
2bro	CHECKPOINT KINASE 1	2.20	Vernalis

2bsm	HEAT SHOCK PROTEIN 90	2.05	Vernalis
2btr	CYCLIN-DEPENDENT KINASE II	1.85	Vernalis
2bts	CYCLIN-DEPENDENT KINASE II	1.99	Vernalis
2byh	HEAT SHOCK PROTEIN 90	1.90	Vernalis
2byi	HEAT SHOCK PROTEIN 90	1.60	Vernalis
2bz5	HEAT SHOCK PROTEIN 90	1.90	Vernalis
2c6e	AURORA KINASE A	2.10	Vernalis
2ccs	HEAT SHOCK PROTEIN 90	1.79	Vernalis
2cct	HEAT SHOCK PROTEIN 90	2.30	Vernalis
2exm	CYCLIN-DEPENDENT KINASE II	1.80	Vernalis
2f4j	ABL KINASE	1.91	Vernalis

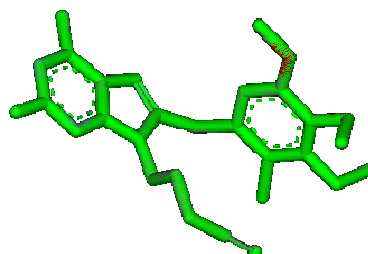
Figure S1. Superposition of the lowest energy conformers obtained by systematic (colored by element) and stochastic (green) searches from MOE (see section 3.2), to illustrate their structural similarities measured by pairwise RMSDs.

Examples are provided for a range of RMSDs. The majority of the compounds (91.4 %) fall in the $\text{RMSD} \leq 0.25\text{\AA}$ category. The PDB entries are given.

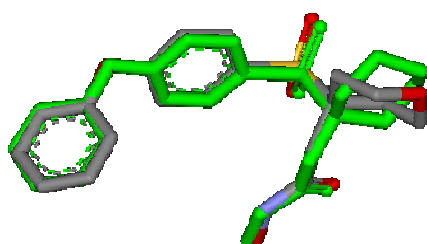
(A) 2AYR, $\text{RMSD} \leq 0.25\text{\AA}$



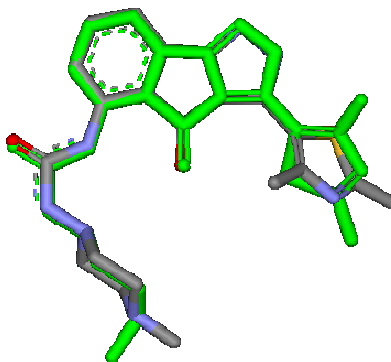
(B) 1UYF, $\text{RMSD} \leq 0.25\text{\AA}$



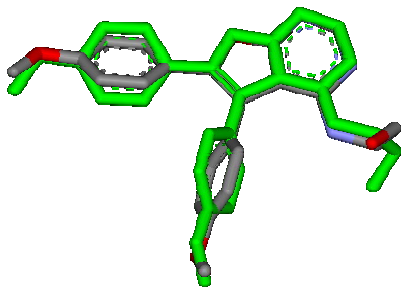
(C) 966C, $\text{RMSD} \sim 0.5\text{\AA}$



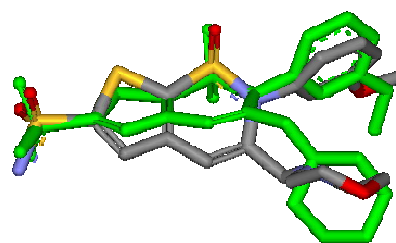
(D) 2B52, $\text{RMSD} \sim 1.0\text{\AA}$



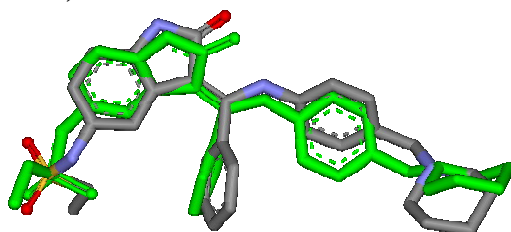
(E) 2BR1, $\text{RMSD} \sim 1.25\text{\AA}$



(F) 1I8Z, $\text{RMSD} \sim 1.50\text{\AA}$



(G) 2BFY, $\text{RMSD} \sim 1.75\text{\AA}$



(H) 1URW, $\text{RMSD} \sim 2.0\text{\AA}$

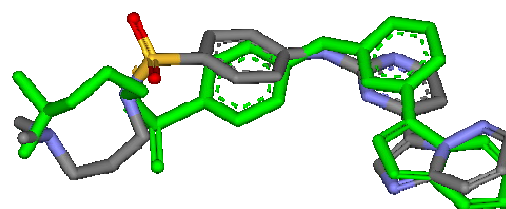
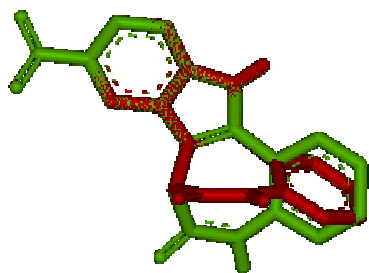
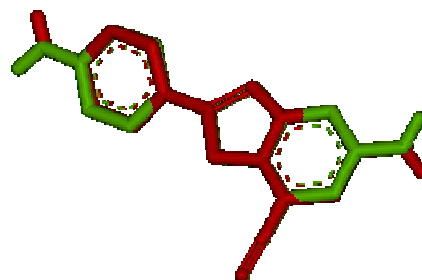


Figure S2. Examples of compounds with the lowest (A) and highest (B) values of percentage of conformational extension ConfExtent. Values of ConfExtent calculated with the solvent accessible surface area (ASA) and the radius of gyration (Rgyr) are given, together with the PDB codes. The most compact conformers in red are overlaid on their extended counterpart in green.

A)

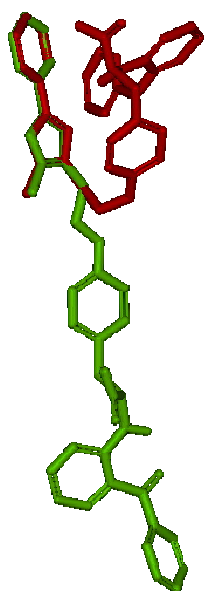


PDB entry 1Q3W
ConfExtent from ASA = 0%
ConfExtent from Rgyr = 0%

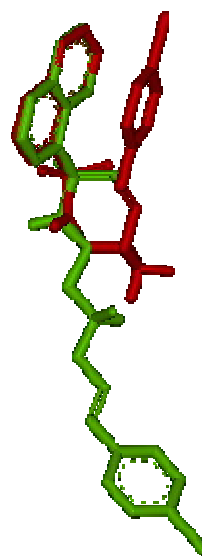


PDB entry 1X76
ConfExtent from ASA = 2%
ConfExtent from Rgyr = 0%

B)



PDB entry 1FM9
ConfExtent from ASA = 38%
ConfExtent from Rgyr = 101%



PDB entry 1YDT
ConfExtent from ASA = 31%
ConfExtent from Rgyr = 109%