# Supporting Information

## Pharmacophore modeling using Site-Identification by Ligand Competitive Saturation (SILCS) with multiple probe molecules

*Wenbo Yu, Sirish Kaushik Lakkaraju, E. Prabhu Raman, Lei Fang and Alexander D. MacKerell, Jr.**

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201

*Correspondence to alex@outerbanks.umaryland.edu

**Table S1**. Overlap coefficients calculated from SILCS simulations 1-5 and 6-10 for generic FragMaps for all eight targets.

| | APOLAR | HBDON | HBACC | POS | NEG |
|---|---|---|---|---|---|
| HIVPR | 0.84 | 0.79 | 0.79 | 0.79 | 0.75 |
| FXa | 0.85 | 0.79 | 0.80 | 0.80 | 0.74 |
| DHFR | 0.84 | 0.78 | 0.78 | 0.77 | 0.70 |
| FGFr1 | 0.83 | 0.78 | 0.79 | 0.78 | 0.71 |
| P38 MAP | 0.82 | 0.79 | 0.79 | 0.78 | 0.71 |
| ADA | 0.82 | 0.78 | 0.79 | 0.77 | 0.73 |
| $ER_{antagonist}$ | 0.79 | 0.78 | 0.78 | 0.77 | 0.71 |
| AmpC | 0.83 | 0.78 | 0.79 | 0.79 | 0.73 |

**Table S2**. GFE cutoffs used for FragMaps to develop SILCS pharmacophore features for all eight targets and the number of features identified for the target.

| | APOLAR | HBDONp | HBACC | POSp | NEG | Number of features |
|---|---|---|---|---|---|---|
| HIVPR | -0.80 | -1.15 | -1.30 | -3.80 | -3.80 | 6 |
| FXa | -1.50 | -1.60 | -1.60 | -2.00 | -2.00 | 5 |
| DHFR | -1.10 | -1.05 | -1.60 | -2.50 | -3.10 | 6 |
| FGFr1 | -1.60 | -1.70 | -1.70 | -2.10 | -2.10 | 5 |
| P38 MAP | -1.30 | -1.60 | -1.60 | -2.30 | -2.30 | 6 |
| ADA | -1.30 | -0.90 | -1.20 | -2.05 | -2.05 | 6 |
| $ER_{antagonist}$ | -1.50 | -0.80 | -1.60 | -1.80 | -1.80 | 4 |
| AmpC | -1.60 | -1.10 | -1.00 | -2.50 | -2.50 | 4 |

**Table S3**. Enrichment factors and AUCs for all pharmacophore models that were generated by SILCS-Pharm and tested against the DUD database for the eight targets.

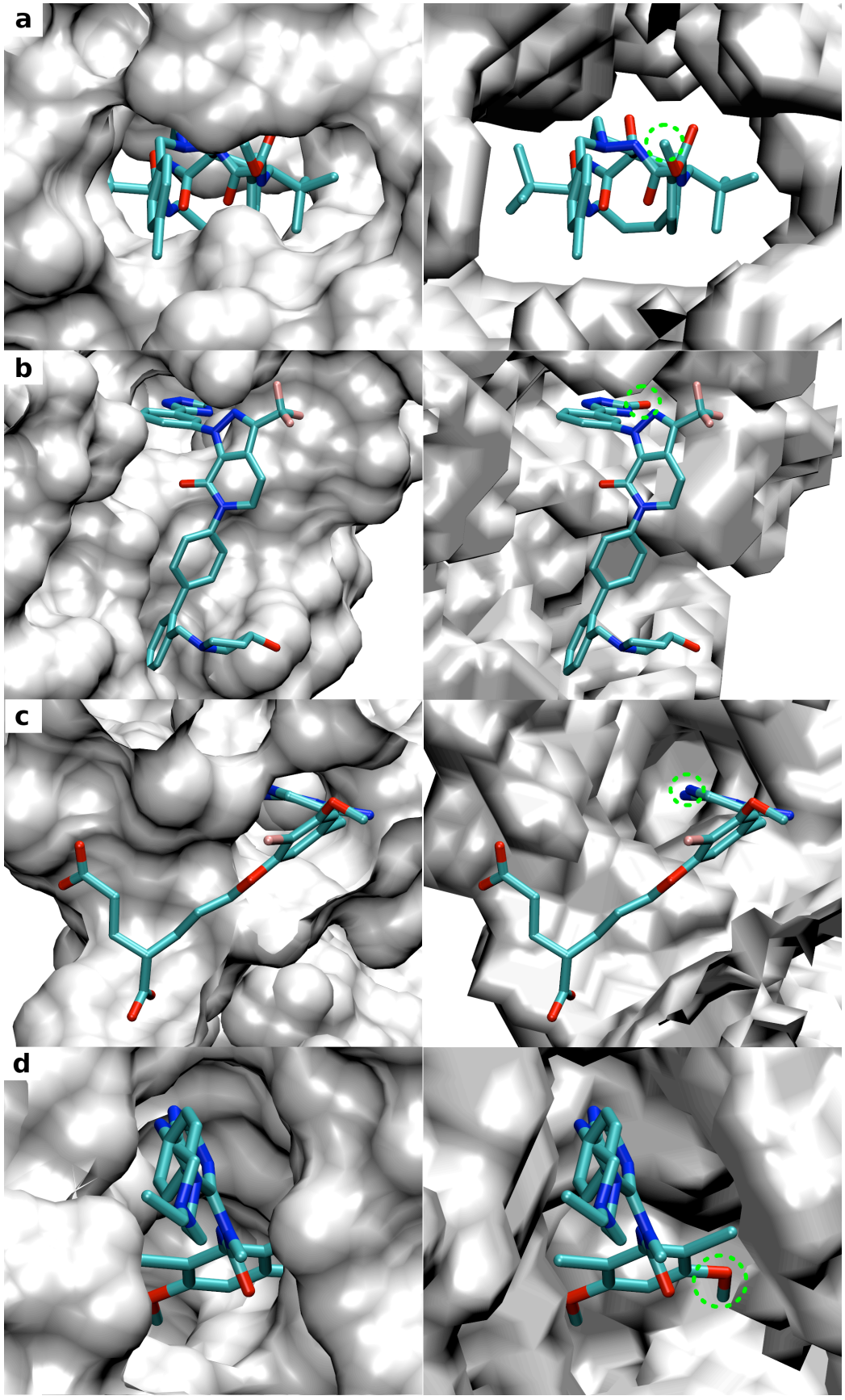
| Targets | Number of key features | $EF_1$ | $EF_{10}$ | $EF_{20}$ | AUC |
|---|---|---|---|---|---|
| HIVPR | 3 | 0.0 | 1.3 | 1.7 | 0.64 |
| | 4 | 6.7 | 4.1 | 3.6 | 0.78 |
| | 5 | 4.2 | 0.4 | 0.2 | 0.05 |
| | 6 | 4.2 | 0.4 | 0.2 | 0.05 |
| FXa | 3 | 4.7 | 11.0 | 5.1 | 0.89 |
| | 4 | 41.9 | 11.4 | 5.1 | 0.91 |
| | 5 | 2.2 | 0.3 | 0.2 | 0.05 |
| DHFR | 3 | 29.3 | 4.8 | 3.0 | 0.79 |
| | 4 | 3.8 | 3.6 | 2.5 | 0.76 |
| | 5 | 5.0 | 0.4 | 0.2 | 0.06 |
| | 6 | 0.5 | 0.0 | 0.0 | 0.02 |
| FGFr1 | 3 | 0.8 | 1.5 | 0.8 | 0.55 |
| | 4 | 0.0 | 0.1 | 0.7 | 0.24 |
| | 5 | 0.0 | 0.0 | 0.0 | 0.01 |
| P38 MAP | 3 | 6.4 | 1.8 | 1.3 | 0.57 |
| | 4 | 0.0 | 1.3 | 1.1 | 0.46 |
| | 5 | 0.4 | 0.5 | 0.6 | 0.29 |
| | 6 | 1.2 | 0.6 | 0.5 | 0.11 |
| ADA | 3 | 0.0 | 3.8 | 3.0 | 0.80 |
| | 4 | 0.0 | 5.4 | 2.7 | 0.65 |
| | 5 | 11.9 | 0.9 | 0.4 | 0.10 |
| | 6 | 0.0 | 0.0 | 0.0 | 0.00 |
| $ER_{antagonist}$ | 3 | 0.0 | 0.0 | 0.1 | 0.49 |
| | 4 | 14.3 | 10.3 | 4.6 | 0.81 |
| AmpC | 3 | 5.0 | 2.0 | 3.6 | 0.70 |
| | 4 | 0.0 | 1.0 | 1.2 | 0.28 |

**Table S4**. Atom classification rules used to translate CGenFF atom types to SILCS FragMap types for LGFE scoring.

| FragMap type[a] | Atom type | Bonded atom type[b] | FragMap type[a] | Atom type | Bonded atom type[b] | FragMap type[a] | Atom type | Bonded atom type[b] |
|---|---|---|---|---|---|---|---|---|
| NCLA | HGA1 | - | AROM | CG2R71 | - | GEND | NG3C51 | - |
| NCLA | HGA2 | - | AROM | CG2RC7 | - | NCLA | NG3N1 | - |
| NCLA | HGA3 | - | ALIP | CG301 | - | FORN | NG3N1 | HGP1 |
| NCLA | HGA4 | - | ALIP | CG302 | - | MAMN | NG3P0 | - |
| NCLA | HGA5 | - | ALIP | CG311 | - | MAMN | NG3P1 | - |
| NCLA | HGA6 | - | NCLA | CG312 | - | MAMN | NG3P2 | - |
| NCLA | HGA7 | - | NCLA | CG314 | - | MAMN | NG3P3 | - |
| NCLA | HGAAM | - | ALIP | CG321 | - | AALO | OG2D1 | - |
| NCLA | HGAAM | - | NCLA | CG322 | - | FORO | OG2D1 | NG2S0 |
| NCLA | HGAAM | - | NCLA | CG323 | - | FORO | OG2D1 | NG2S1 |
| NCLA | HGP1 | - | NCLA | CG324 | - | FORO | OG2D1 | NG2S2 |
| NCLA | HGP2 | - | ALIP | CG331 | - | ACEO | OG2D2 | - |
| NCLA | HGP3 | - | NCLA | CG334 | - | AALO | OG2D3 | - |
| NCLA | HGP4 | - | NCLA | CG3AM0 | - | AALO | OG2D4 | - |
| NCLA | HGP5 | - | NCLA | CG3AM1 | - | FORO | OG2D4 | NG2R61 |
| NCLA | HGPAM | - | NCLA | CG3AM2 | - | AALO | OG2D5 | - |
| NCLA | HGPAM | - | ALIP | CG3C31 | - | AALO | OG2N1 | - |
| NCLA | HGPAM | - | ALIP | CG3C41 | - | AALO | OG2P1[c] | - |
| NCLA | HGR51 | - | ALIP | CG3C50 | - | ACEO | OG2P1[c] | - |
| NCLA | HGR52 | - | ALIP | CG3C51 | - | GENA | OG2R50 | - |
| NCLA | HGR53 | - | ALIP | CG3C52 | - | GENA | OG3R60 | - |
| NCLA | HGR61 | - | NCLA | CG3C53 | - | GENA | OG301 | - |
| NCLA | HGR62 | - | NCLA | CG3C54 | - | GENA | OG302 | - |
| NCLA | HGR63 | - | ALIP | CG3RC1 | - | ACEO | OG303 | - |
| NCLA | HGR71 | - | GENA | NG1T1 | - | ACEO | OG304 | - |
| GENN | CG1T1 | - | GENA | NG2D1 | - | MEOO | OG311 | - |
| NCLA | CG1N1 | - | NCLA | NG2S0 | - | ACEO | OG312 | - |
| GENN | CG2D1 | - | FORN | NG2S1 | - | GENA | OG3C51 | - |
| GENN | CG2D2 | - | FORN | NG2S2 | - | GENA | OG3C61 | - |
| NCLA | CG2D1O | - | FORN | NG2S3 | - | AALO | SG2D1 | - |
| NCLA | CG2D2O | - | NCLA | NG2O1 | - | GENA | SG2R50 | - |
| AROM | CG2DC1 | - | MAMN | NG2P1 | - | MEOO | SG311 | - |
| AROM | CG2DC2 | - | FORN | NG2R43 | - | GENA | SG301 | - |
| AROM | CG2DC3 | - | GENA | NG2R50 | - | ACEO | SG302 | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NCLA | CG2N1 | - | GENA | NG2R51 | - | NCLA | SG3O1 | - |
| NCLA | CG2N2 | - | NCLA | NG2R51 | CG3C51 | NCLA | SG3O2 | - |
| NCLA | CG2O1 | - | NCLA | NG2R51 | CG3C52 | NCLA | SG3O3 | - |
| GEND | CG2O1 | HGR52 | NCLA | NG2R51 | CG3C53 | GENA | FGA1 | - |
| GEND | CG2O2 | HGR52 | NCLA | NG2R51 | CG331 | GENA | FGA2 | - |
| NCLA | CG2O2 | - | NCLA | NG2R51 | CG321 | NCLA | FGA3 | - |
| NCLA | CG2O3 | - | NCLA | NG2R51 | CG311 | ACEO | FGP1 | - |
| NCLA | CG2O4 | - | GEND | NG2R51 | HGP1 | GENA | FGR1 | - |
| NCLA | CG2O5 | - | NCLA | NG2R52 | - | ALIP | CLGA1 | - |
| NCLA | CG2O6 | - | GEND | NG2R52 | HGP1 | ALIP | CLGA3 | - |
| NCLA | CG2O7 | - | MAMN | NG2R52 | HGP2 | ALIP | CLGR1 | - |
| GENN | CG2R51 | - | NCLA | NG2R53 | - | ALIP | BRGA1 | - |
| NCLA | CG2R52 | - | FORN | NG2R53 | HGP1 | ALIP | BRGA2 | - |
| NCLA | CG2R53 | - | GENA | NG2R60 | - | ALIP | BRGA3 | - |
| MAMN | CG2R53 | HGR53 | NCLA | NG2R61 | - | ALIP | BRGR1 | - |
| AROM | CG2R61 | - | GEND | NG2R61 | HGP1 | ALIP | IGR1 | - |
| AROM | CG2R62 | - | GENA | NG2R62 | - | NCLA | PG0 | - |
| NCLA | CG2R63 | - | NCLA | NG2RC0 | - | NCLA | PG1 | - |
| NCLA | CG2R64 | - | GENA | NG301 | - | NCLA | PG2 | - |
| NCLA | CG2R66 | - | GEND | NG311 | - | NCLA | ALG1 | - |
| AROM | CG2R67 | - | GEND | NG321 | - | | | |
| AROM | CG2RC0 | - | GEND | NG331 | - | | | |

[a] SILCS FragMap types: NCLA, non-classified (zero occupancy map); AROM, aromatic (benzene carbons); ALIP, aliphatic (propane carbons); GENN, generic nonpolar (benzene and propane carbons); GEND, generic hydrogen bond donor (formamide nitrogen); GENA, generic hydrogen bond acceptor (formamide and acetaldehyde oxygens); MEOO, methanol oxygen; FORN, formamide nitrogen; FORO, formamide oxygen; MAMN, positive methylammonium nitrogen; ACEO, negative acetate oxygens; AALO, acetaldehyde oxygen. [b] Bonded atoms are used to determine specific cases where the assignment is done considering atom bonding environment. Such specific rules for an atom are applied over the general rule for the same atom. For example, for CGenFF atom type OG2D4, it is assigned FORO type if it is covalently bonded to an atom with NG2R61 type, otherwise it is assigned AALO FragMap type. [c] For OG2P1, a charge criterion is also applied. If the partial atomic charge for an OG2P1 type atom is less than -0.55, it is classified as ACEO FragMap type, otherwise, AALO type is assigned.


a
b
c
d

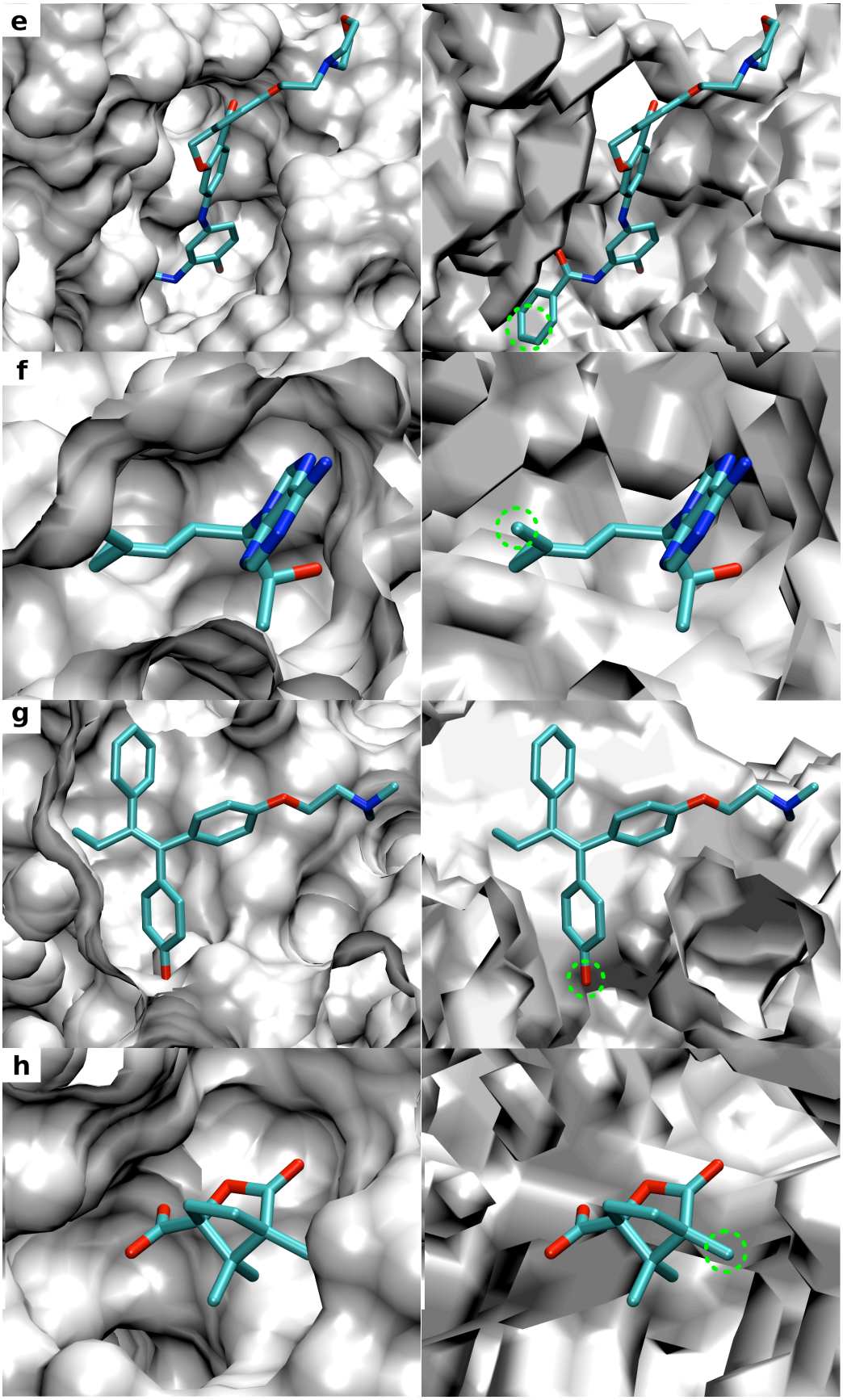



Figure S1. Comparison of the solvent accessible surface of crystal protein structure used to initialize the SILCS simulation (left panel) and SILCS exclusion map (right panel) for each target: (a) HIVPR (PDB 1G2K); (b) FXa (PDB

1FJS); (c) DHFR (PDB 3DFR); (d) FGFr1 (PDB 3KY2); (e) P38 MAP (PDB 1OUY); (f) ADA (PDB 1NDW); (g) ER (PDB 3ERT); (h) AmpC (PDB 1XGJ). The crystal binding orientation of a selected ligand for each target presents in protein-ligand complex other than the one used for SILCS simulation is also shown: (a) HIVPR (PDB 3ZPS); (b) FXa (PDB 3FFG); (c) DHFR (PDB 1DIU); (d) FGFr1 (PDB 3TT0); (e) P38 MAP (PDB 4L8M); (f) ADA (PDB 2Z7G); (g) ER (PDB 2BJ4); (h) AmpC (PDB 4KZ7). The green dashed circle indicates ligand atoms that have clashes with the protein surface but not the exclusion map.