Supporting Information

Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics

S. Roy Kimura,[*,#] *Hai Peng Hu,*[±] *Anatoly Ruvinsky,*[§] *Woody Sherman,*[§,¶] *and Angelo D Favia*[*,±,^]

[#]Schrödinger KK, 17th Fl, Marunouchi Trust Tower North, 1-8-1 Marunouchi, Chiyoda-ku, Tokyo, Japan

[§]Schrödinger LLC, 222 Third Street, Suite 2230, Cambridge, MA 02142

[±]Lilly China Research and Development Center (LCRDC), Eli Lilly and Company, Building 8, 338 Jia Li Lue Road, Shanghai 201203, PR China

[¶]Current Address: Silicon Therapeutics, 300 A St., Boston, MA 02210

[^]Current Address: Elanco, 2500 Innovation way, Greenfield, IN 46140

*ADF, E-mail: favia_angelo@lilly.com, ad.favia@gmail.com

*SRK, E-mail: leadingelement@gmail.com

ASSOCIATED CONTENT

Supporting Information

1. RMSD trajectory plots

2. All PCA density distribution plots

3. PCA analysis of the relative side chain positions at the cryptic sites for 2 independent simulations run for Kinesin Eg5

4. Analysis of the biological relevance of the top 6 ranked hotspots identified via mixed-solvent Simulations

5. RMSD analysis of the heavy atoms and Cα of the cryptic site

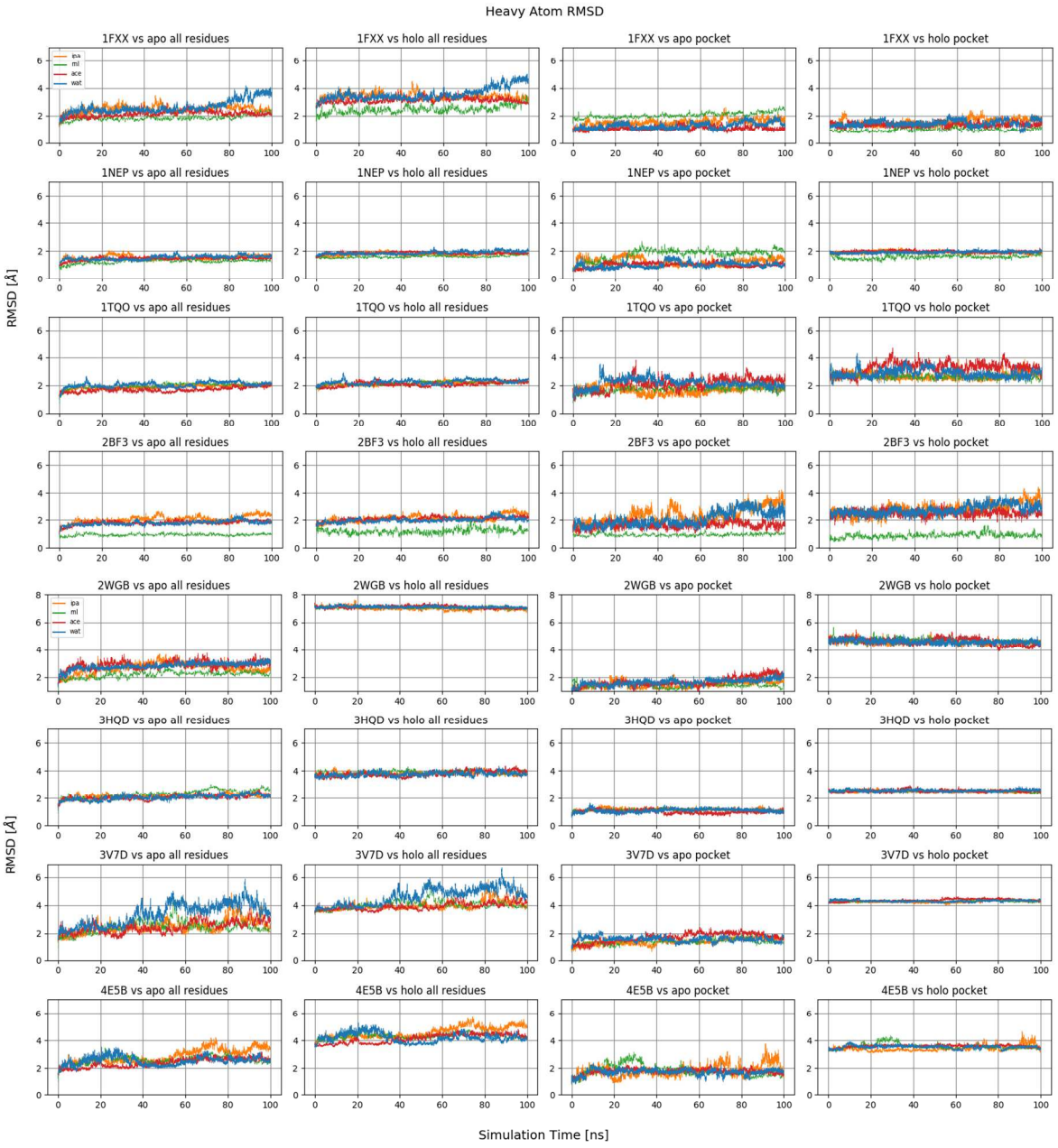

Figure S1. Root mean square deviation of all protein heavy atoms and protein heavy atoms near the cryptic sites with respect to the apo and holo crystal structures.

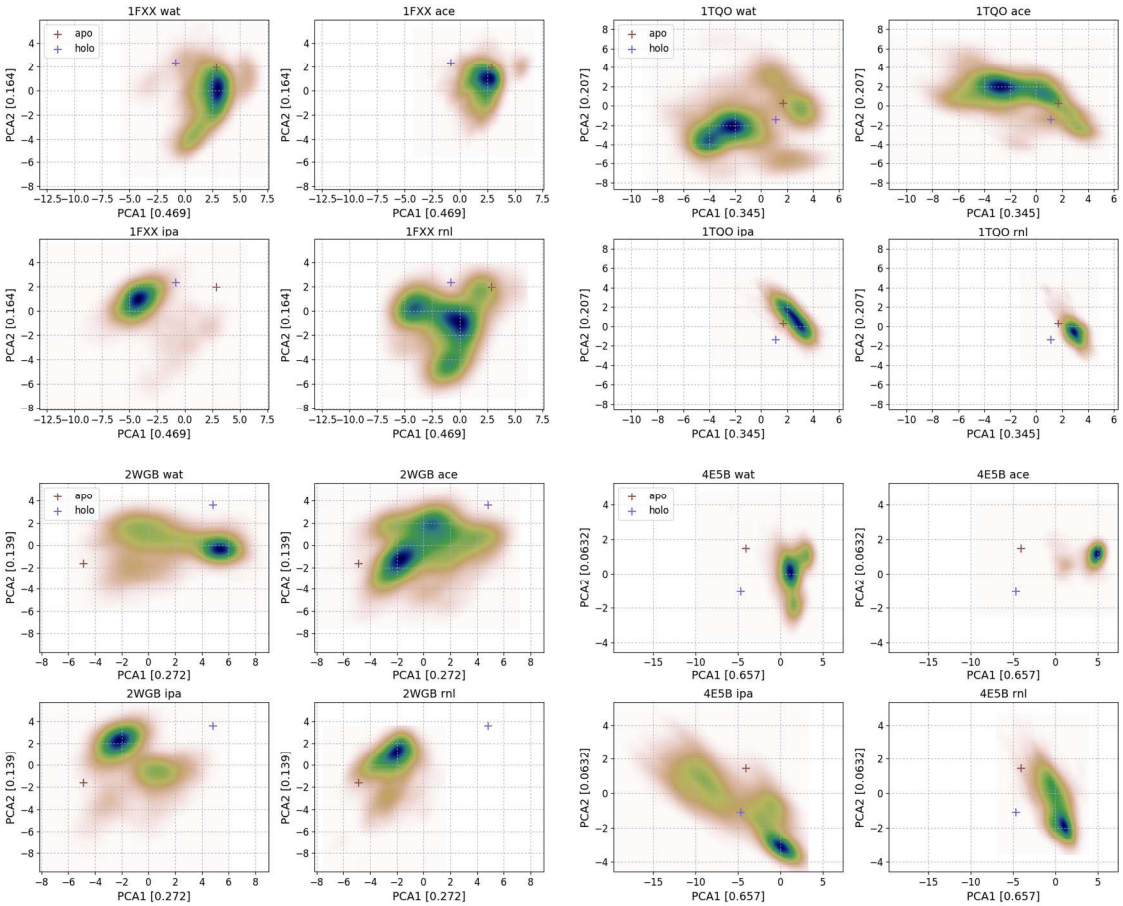

Figure S2. PCA analysis of the relative side chain positions at the cryptic sites. Shown are density distributions of the MD frames using different mixed solvent probes, plotted against the first 2 PCA vectors. See Results and Methods sections for details.

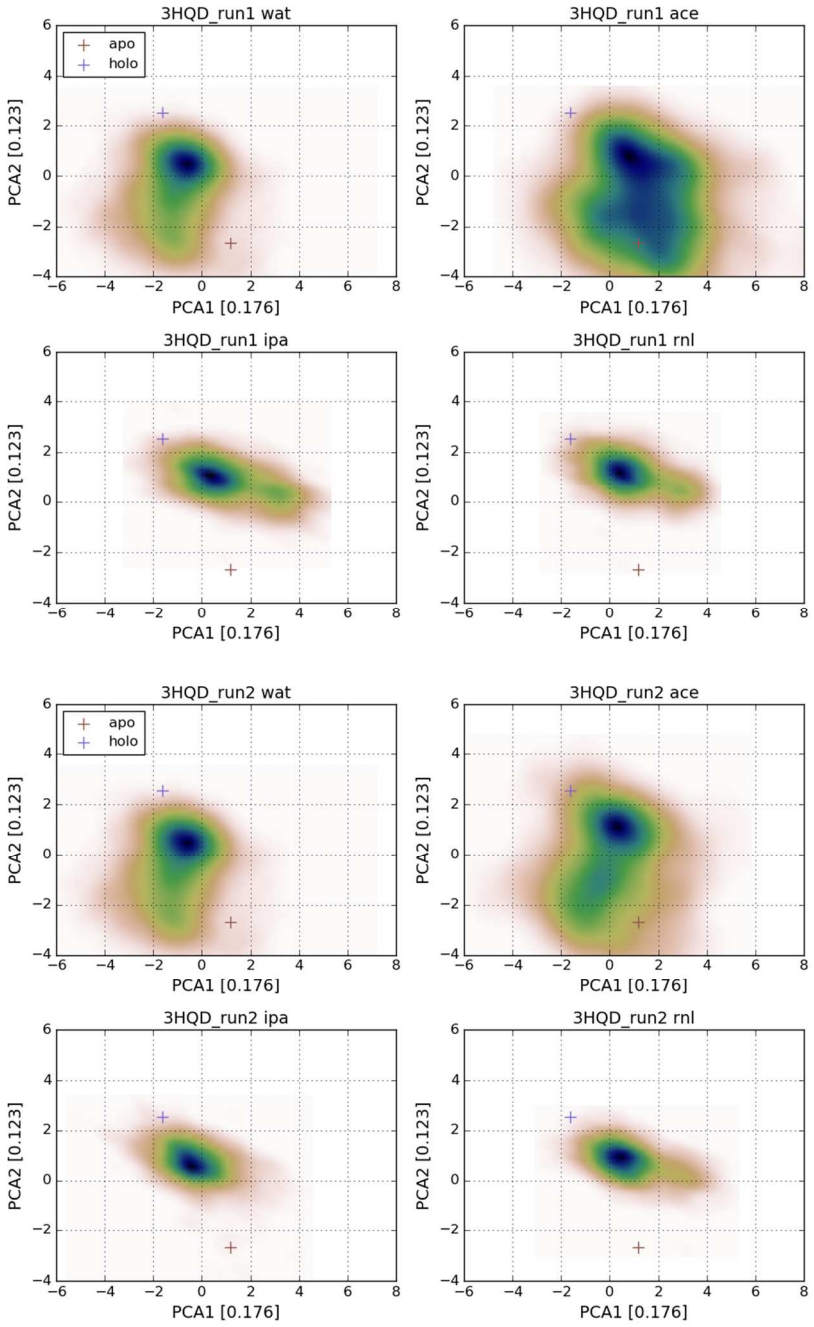

Figure S3. PCA analysis of the relative side chain positions at the cryptic sites for 2 independent simulations run for Kinesin Eg5. The water simulation was used as a reference for both set of plots. Shown are density distributions of the MD frames using different mixed solvent probes, plotted against the first 2 PCA vectors. See Results and Methods sections for details

Table S1. Analysis of the biological relevance of the top 6 ranked hotspots identified via Mixed Solvent Simulations (C.S. = Cryptic Site; A.S. = Additional Known Binding Site; C.F. = Complex Formation Site; N. = New Site). Note that a) the Additional Known Binding Site could be either the orthosteric site or a known allosteric site, b) the Cryptic Site can be either orthosteric or allosteric.

Rank# / System	1	2	3	4	5	6
Exonuclease I*	A.S.	A.S.	N.	A.S	N.	C.S.
Niemann-Pick C2 Protein	N.	N.	C.S.	N.	N.	N.
Staphylococcal Nuclease	C.S.	N.	N.	C.F.	N.	N.
Toluene-4-Monooxygenase	C.F.	C.F.	C.S.	N.	N.	N.
TETR-Like Transcriptional Regulator LFRR	C.F.	C.F.	C.S.	N.	N.	N.
Kinesin Eg5	N.	C.F.	C.F.	C.S.	N.	N.
Cdc4	A.S.	C.S.	N.	N.	N.	N.
P38α	C.S.	A.S.	N.	N.	N.	N.

Table S2. Pocket heavy atom and Cα RMSD values of the Apo and mixed solvent induced structures with respect to the holo reference.

System	Pocket Residue Numbers	Pocket Heavy Atom RMSD		Pocket Cα RMSD	
		Apo	MSS	Apo	MSS
1FXX	245, 312, 313, 317, 327, 331	1.43	1.60	0.73	0.85
1NEP	64, 66, 100, 101	2.28	1.80	1.80	1.17
1TQ0	87, 113, 115	4.41	3.52	2.13	1.68
2BF3	75, 76,78, 82, 88, 95	2.59	1.51	1.14	1.06
2WGB	67, 71, 124, 126	2.59	3.23	2.33	2.36
3HDQ	116, 127, 211	2.44	2.13	1.06	1.38
3V7D	629, 630, 631, 664	2.81	3.41	1.52	1.62