POVME 3.0: Software for Mapping Binding Pocket Flexibility

Supporting Information.

User Notes and Best Practices

Effect of structure alignment on pocket analysis

In the course of performing this work, we determined that robust alignment of the protein pocket is a prerequisite for successful POVME analysis. Many trajectory-handling programs such as VMD can perform RMSD alignments, and most default to alignments based on the entire protein (*e.g.,* all alpha carbons). However, some proteins undergo significant domain motion, so care should be taken to perform the alignment such that the binding pocket remains in the same location in Cartesian space. This may require performing an alignment of only the domain containing the binding pocket, or restricting the alignment to a set of pocket-lining residues. During the development and testing of POVME 3.0, inappropriate alignment of the protein trajectory/ensemble was a common problem. Misalignment is usually noticed during clustering and PCA, and is represented by difference regions that line surfaces on opposite sides of the binding pocket. In these cases, one entire face of the pocket is seen to lose volume over its surface, and the opposite face is seen to gain it. This type of change is likely an artifact, adding noise to the interpretation of pocket dynamics. As a solution, we investigated the possibility of providing tools for alignment of pocket shapes, but POVME's voxel-based representation was found to be poorly suited for this task.

We advise against interpretation of scalar volume values

The value that POVME provides for pocket volume is simply the sum of the volumes of the voxels comprising the pocket. Because heuristics are used to define the outer boundary of the pocket, the numerical value of the pocket volume is difficult to meaningfully compare between programs, or even significantly different pockets analyzed by the same program. Users should take caution when comparing the POVME-provided volumes to anything except highly similar pockets. Without knowing how users plan to interpret or compare volume numbers we cannot ensure that they are fit for a specific purpose. Instead, we encourage users to compare pockets in 3D. POVME provides directly visualizable outputs as well as Python functions for performing mathematical operations on sets of pocket shapes. Given the frame-by-frame output files provided by POVME, users with Python knowledge can load the sets of pockets as lists of points, then use POVME functions to compute their difference and output it as a pdb or dx file for visualization.

Inclusion and seed regions must be identically defined for successful post-analysis

As the clustering and PCA processes consider variation in pocket shape, it is important that the volume eligible to be part of the pocket is consistently defined for all frames being studied. In other words, post-processing requires that the inclusion and seed regions be identically defined for all of the trajectories being analyzed. The provided workflows take care of this step automatically, by taking as input a user-defined inclusion and seed region. However, when running POVME analysis separately on multiple trajectories with the intent of combining their results in post-analysis, it is essential that their inclusion and seed regions be the same.

Post-processing analysis will not work if the pockets being analyzed have different boundaries.

The outer boundary of the pocket is defined both by the edge of the inclusion region, and if the "ConvexHullExclusion" keyword is used, by the convex hull of the protein. When comparing pocket volumes within the same trajectory, users should ensure that the boundary of the pocket is consistently defined.

Recalculating a different convex hull for each frame of a trajectory adds noise to quantitative analysis, as the convex hull definition is sensitive to movements of surface residues. Because many pockets widen as they approach the surface of the protein, small changes in how the outer boundary is defined can lead to large numbers of points being added to or removed from the pocket. During quantitative analysis, this large number of variable points will outweigh the smaller changes corresponding to pocket dynamics and shape change inside the cavity.

To instruct POVME to use a single definition of this outer boundary, users should ensure that the ConvexHullExclusion option is set to a keyword other than "each". The default keyword, "none", is recommended. While this choice may lead to a large number of points being defined outside of the pocket, POVME's clustering and PCA scripts focus on *differences* in pocket shape, thus points that lie outside of the protein and are never removed from the pocket do not affect the results of the analysis.

On the inner barrier of a pocket, users should be mindful of another potential source of noise. When a pocket of interest is near another cavity, the protein atoms will sometimes rearrange during MD to join the two. When this joining occurs, the pocket region defined by POVME can become much larger, thereby adding noise to post-processing. Two options to avoid this situation are: 1) If a ligand is present, use the "DefinePocketByLigand" keyword to define the pocket as the area immediately around the ligand, or 2) carefully define inclusion and seed regions so that the unwanted cavity is not included in the analysis.

Advice for difficult pockets

All pocket definition algorithms that we know of, POVME included, have trouble meaningfully describing very shallow or very narrow pockets. As "binding pocket" is an abstract concept, there is no single definition for where one ends. However, for an algorithm to successfully make conclusions about a region of a protein responsible for binding does not require such a definition. We propose the following techniques to overcome shortcomings in pocket definition:

For very narrow pockets or channels, we recommend that users define an inclusion region which covers the pocket of interest, and also extends outside the protein. Users should also place the seed region exclusively outside of the protein (but in the inclusion region), and disable the convex hull algorithm. It should be noted that the pocket volume (in cubic Angstroms) calculated by this approach are meaningless, though the *fluctuations* in pocket volume are real. For each frame analyzed by this technique, the POVME pocket "definition" will begin outside the protein and grow inwards, until the pocket becomes too narrow to continue. This technique is also useful when analyzing adjacent pockets to find frames in which they are connected.

For shallow pockets, a certain property of clustering and PCA (that only variance in features is analyzed) largely solves the definition problem. Users should define an inclusion region which extends well outside the protein and disable the convex hull. While the volume in cubic Angstroms will again be meaningless in this technique, the changes in pocket shape will reflect the true motions of the protein surface. Therefore, when POVME clustering / PCA is performed on such data, the large constant set of voxels outside of the pocket has no quantitative effect (as it never varies, and is therefore "erased" by mean normalization), while the meaningful changes in pocket shape on the surface define the voxels that are clustered / described in the principal components.

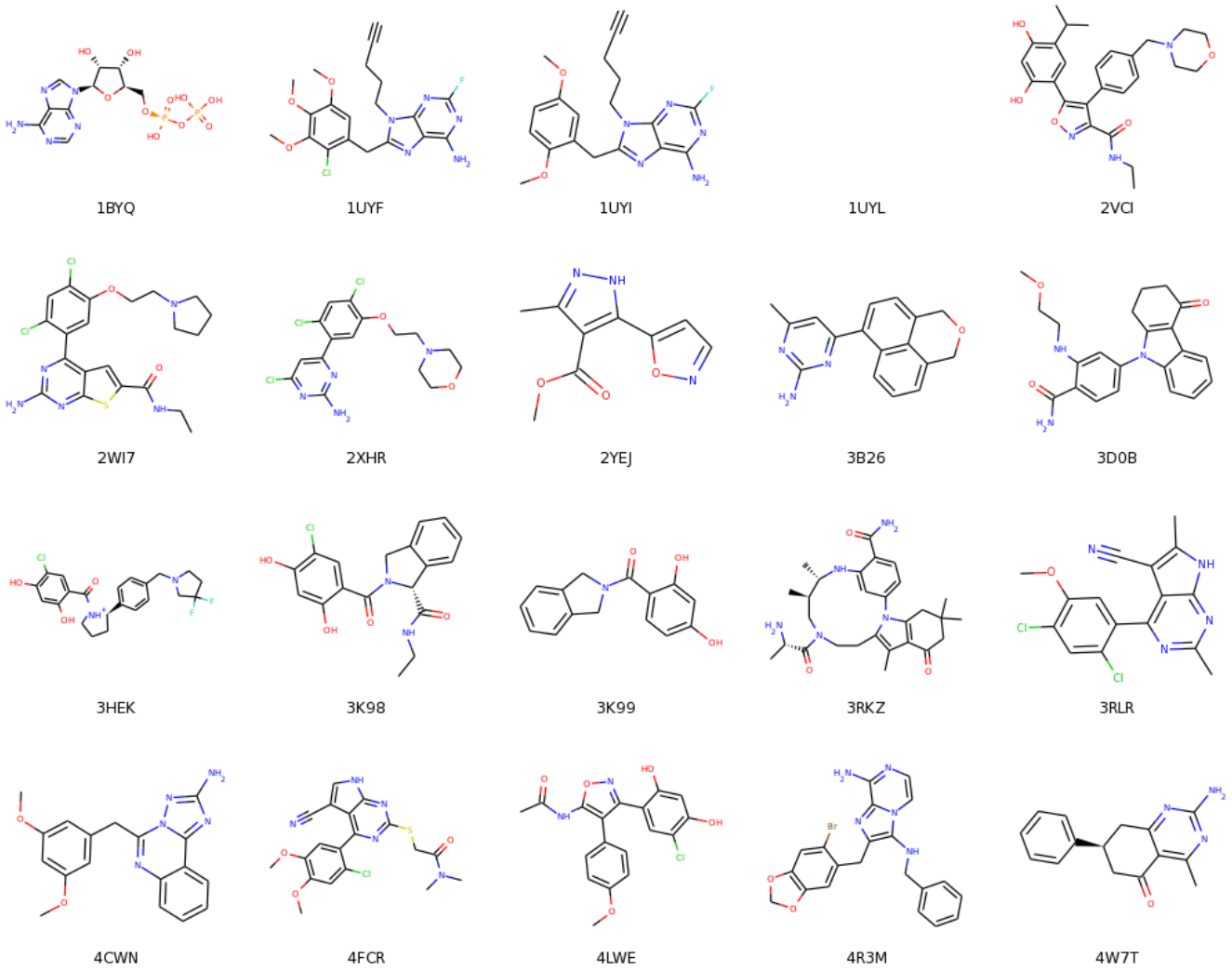


Figure S1: All HSP90 ligands used in this work

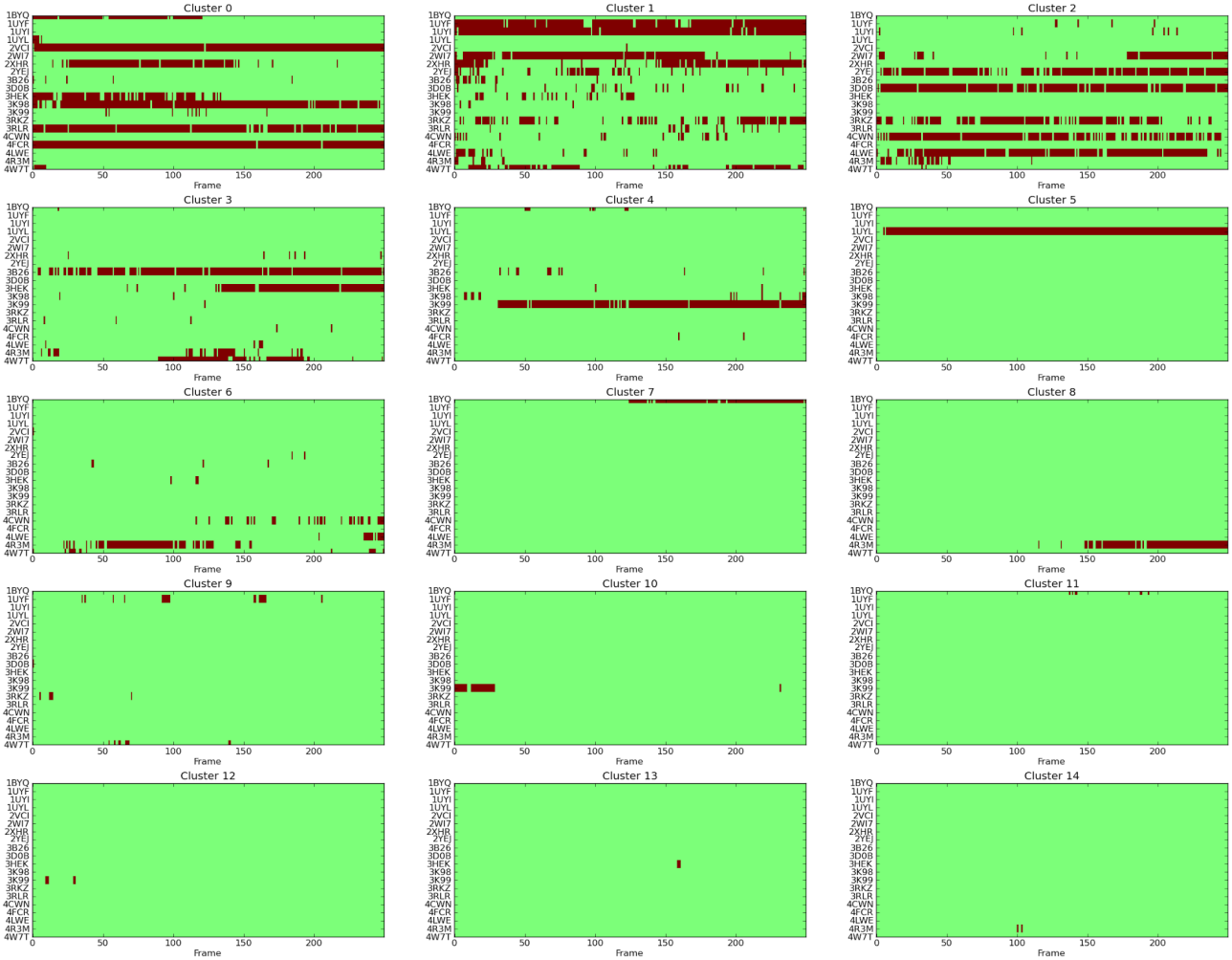


Figure S2: POVME cluster assignments for all frames from HSP90 trajectories.

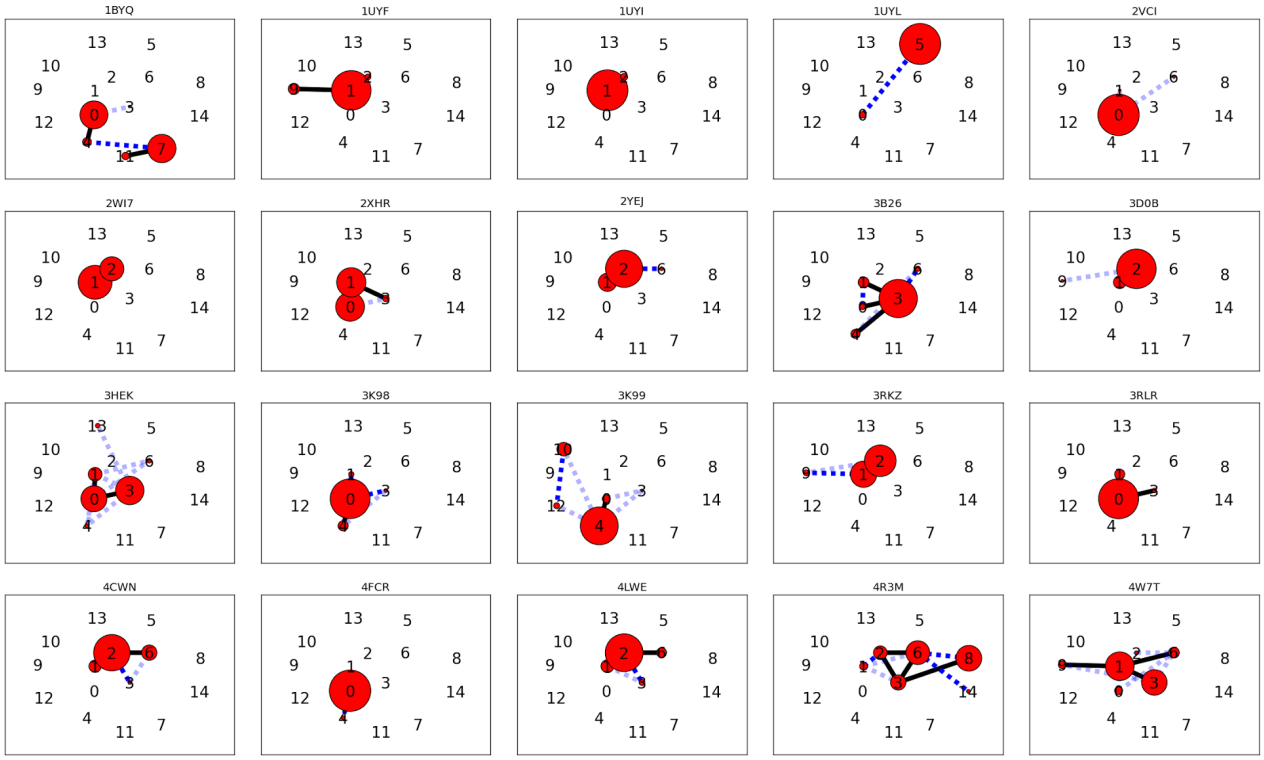


Figure S3: “Kinetic network” depiction of the 20 HSP90 trajectories. Black numbers indicate cluster index (0-14). Red circles indicate the number of frames assigned to each cluster. Edges indicate the number of transitions observed between clusters in the MD trajectories (light blue dashes = 1 or 2, dark blue dashes = 3 to 5, solid black line = greater than 5). Clusters are arranged in 2D according to a force-based layout, in which each pair of clusters is pulled together by a force proportional to the number of observed transitions.

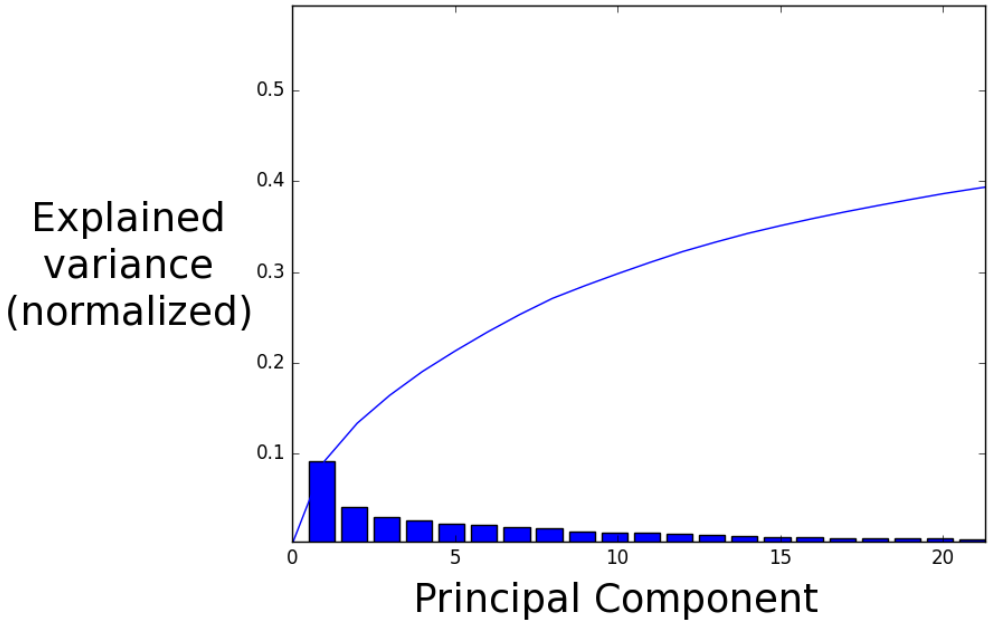
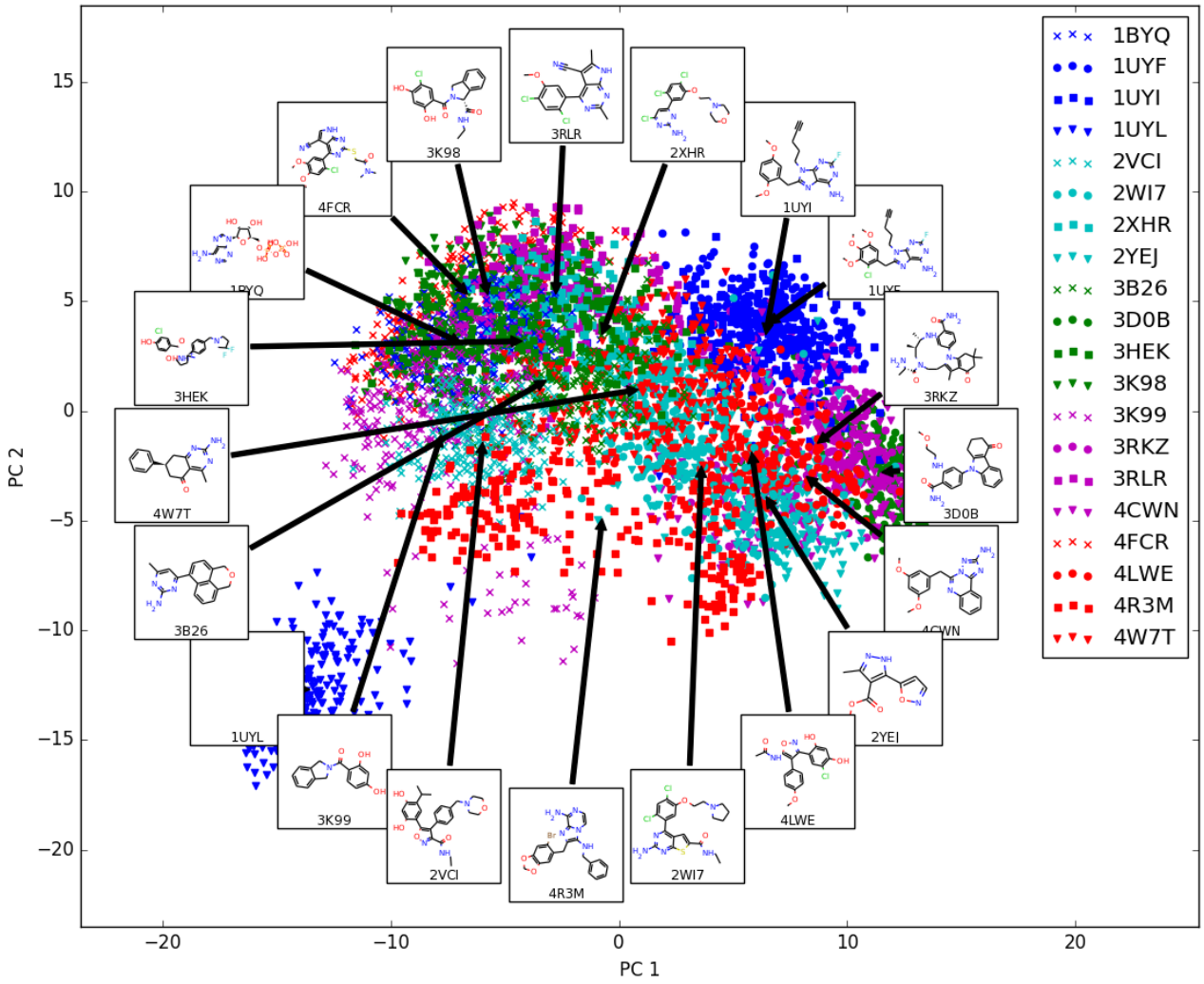
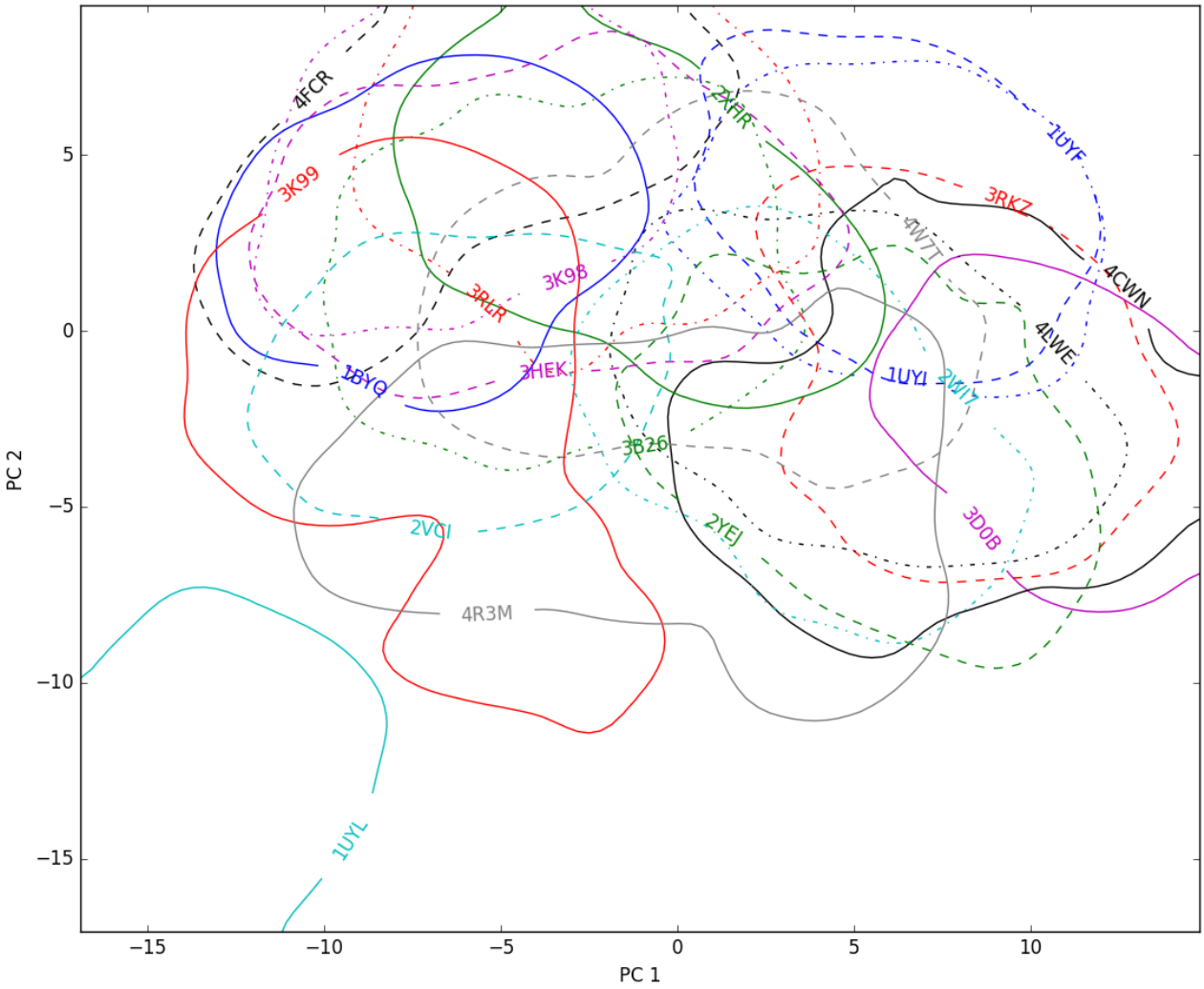


Figure S4: Explained variance plot of Principal Component Analysis of HSP90 trajectories.

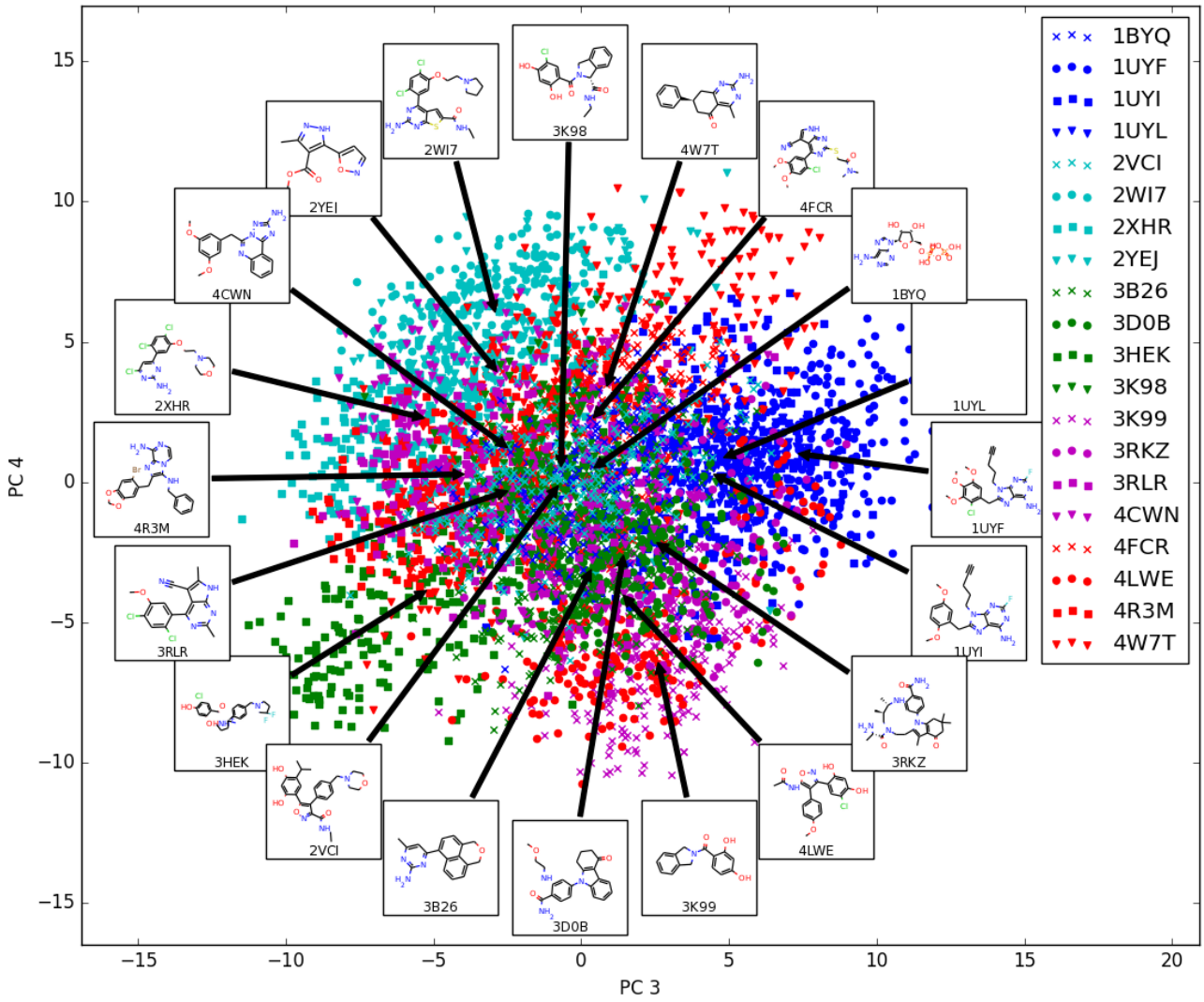
S5A)


PC 1
PC 2
1BYQ
1UYF
1UYI
1UYL
2VCI
2WI7
2XHR
2YEJ
3B26
3D0B
3HEK
3K98
3K99
3RKZ
3RLR
4CWN
4FCR
4LWE
4R3M
4W7T
3K98
3RLR
2XHR
4FCR
1UYI
1BYQ
1UYF
3HEK
3RKZ
4W7T
3D0B
3B26
4CWN
1UYL
2YEJ
3K99
4LWE
2VCI
2WI7
4R3M

S5B)


4FCR
3K99
2XFIR
1UYF
3RKZ
4CWN
4WVT
4LWE
3K98
3RLR
1BYQ
3HEK
1UYI
2WVT
3B26
2YEJ
3D0B
2VCI
4R3M
1UYL
PC 1
PC 2
5
0
−5
−10
−15
−15
−10
−5
0
5
10

S5C)


PC 4
PC 3
15
10
5
0
−5
−10
−15
−15
−10
−5
0
5
10
15
20
1BYQ
1UYF
1UYI
1UYL
2VCI
2WI7
2XHR
2YEJ
3B26
3D0B
3HEK
3K98
3K99
3RKZ
3RLR
4CWN
4FCR
4LWE
4R3M
4W7T
2YEI
2WI7
3K98
4W7T
4FCR
4CWN
1BYQ
2XHR
1UYL
4R3M
1UYF
3RLR
1UYI
3HEK
3RKZ
2VCI
4LWE
3B26
3D0B
3K99

S5D)

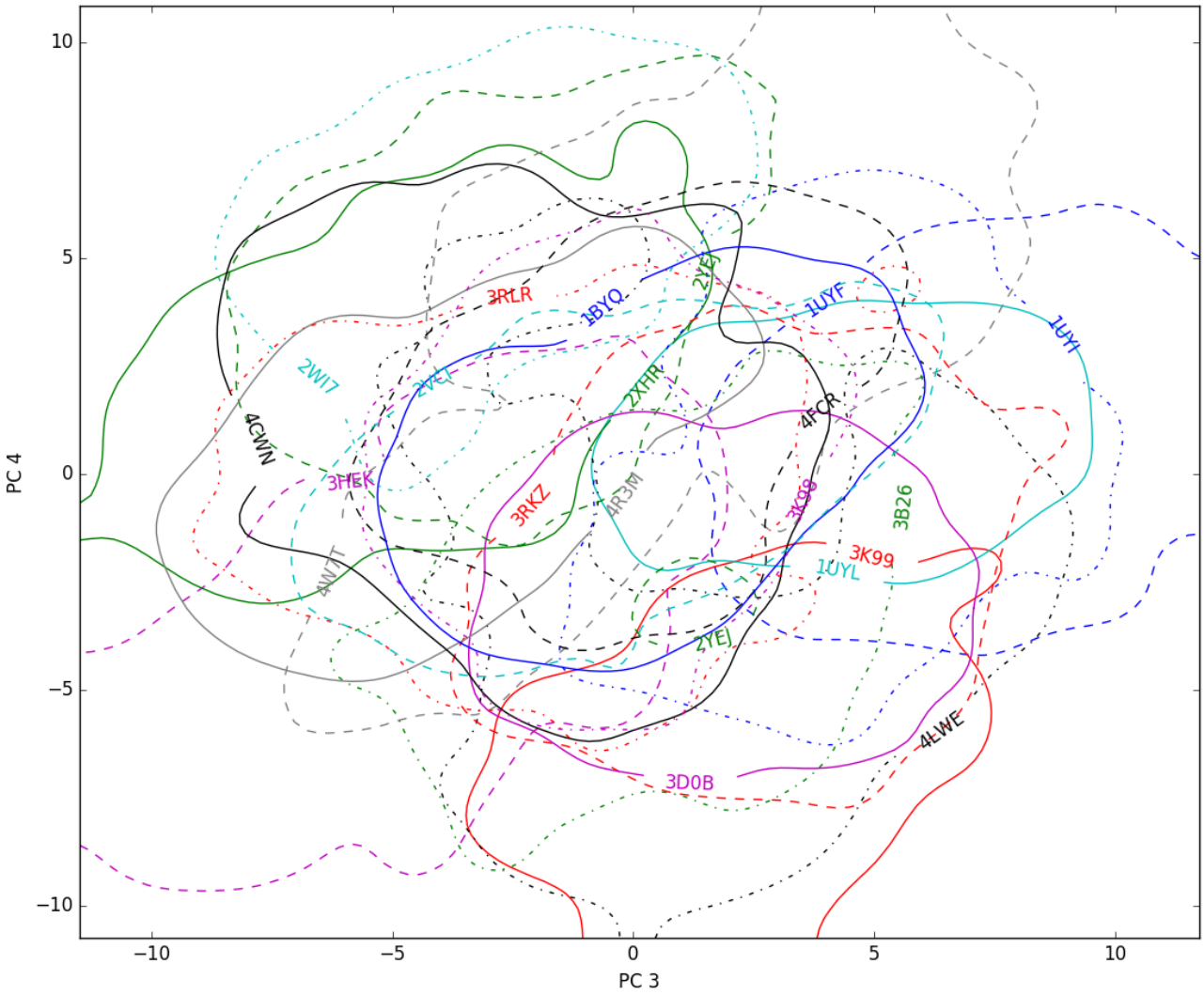


Figure S5: Trajectories in Principal Component Space. A) Evenly-sampled HSP90 frames are overlaid on Principal Components 1 and 2. Each simulation is indicated by a figure of its bound ligand with an arrow pointing to the centroid of its frames in PC space. B) The same data as A, shown as a contour plot. C and D) Same as A and B, for PCs 3 and 4.

$SCHRODINGER/utilities/prepwizard -keepfarwat -disulfides -fillsidechains -fillloops -mse -metal_binding -samplewater -propka_pH 7 -label_pkas ${PDBID} ${PDBID}_prepped.pdb -reference_pdbid 1BYQ -LOCAL

Scheme S1: An example prep command using Schrodinger Protein Prep Wizard

S01-Min01-Proton.in

```
Minimization 01 - Proton
 &cntrl
   imin = 1,            ! Minimization (Yes)
   ntmin = 1,           ! Minimization Method (Steepest descent/Conjugate gradient)
   maxcyc = 2000,       ! Maximum number of minimization cycles (2000 cycles)
   ncyc = 1000,         ! Cycle of switch from steepest descent to conjugate gradient (at cycle 1000)
   cut = 10,            ! Non-bonding Cut-off (10 A)
   ntb = 1,             ! Periodic Conditions (Yes)
   ntr = 1,             ! Harmonic constraints in Cartesian space (Yes)
   restraint_wt = 10.0  ! Positional restraints weight ( 10 kcal/mol-A^2)
   restraintmask = "!@H=",  ! Restrained atoms (Not protons)
 /
```

S02-Min02-Solvent.in

```
Minimization 02 - Solvent
 &cntrl
   imin = 1,               ! Minimization (Yes)
   ntmin = 1,              ! Minimization Method (Steepest descent/Conjugate gradient)
   maxcyc = 2000,          ! Maximum number of minimization cycles (2000 cycles)
   ncyc = 1000,            ! Cycle of switch from steepest descent to conjugate gradient (at cycle
1000)
   cut = 10,               ! Non-bonding Cut-off (10 A)
   ntb = 1,                ! Periodic Conditions (Yes)
   ntr = 1,                ! Harmonic constraints in Cartesian space (On)
   restraint_wt = 10.0,    ! Positional restraints weight ( 10 kcal/mol-A^2)
   restraintmask = ":1-213 & :adp",      ! Restrained atoms (protein and ligand)
 /
```

S03-Min03-Focused.in

```
Minimization 03 - Focused
 &cntrl
   imin = 1,               ! Minimization (Yes)
   ntmin = 1,              ! Minimization Method (Steepest descent/Conjugate gradient)
   maxcyc = 2000,          ! Maximum number of minimization cycles (2000 cycles)
   ncyc = 1000,            ! Cycle of switch from steepest descent to conjugate gradient (at cycle
1000)
   cut = 10,               ! Non-bonding Cut-off (10 A)
   ntb = 1,                ! Periodic Conditions (Yes)
   ntr = 1,                ! Harmonic constraints in Cartesian space (On)
   restraint_wt = 10.0,    ! Positional restraints weight ( 10 kcal/mol-A^2)
   restraintmask = ":1-213",       ! Restrained atoms (protein)
 /
```

S04-Min04-Sidechains.in

```
Minimization 04 - Sidechains and Solvent
 &cntrl
```

```
   imin = 1,                ! Minimization (Yes)
   ntmin = 1,               ! Minimization Method (Steepest descent/Conjugate gradient)
   maxcyc = 2000,           ! Maximum number of minimization cycles (2000 cycles)
   ncyc = 1000,             ! Cycle of switch from steepest descent to conjugate gradient (at cycle
1000)
   cut = 10,                ! Non-bonding Cut-off (10 A)
   ntb = 1,                 ! Periodic Conditions (Yes)
   ntr = 1,                 ! Harmonic constraints in Cartesian space (On)
   restraint_wt = 10.0,     ! Positional restraints weight ( 10 kcal/mol-A^2)
   restraintmask = ":1-213@CA,N,C,O",     ! Restrained atoms (protein backbone)
 /
```

S05-Min05-All.in

```
Minimization 05 - All Atoms
 &cntrl
   imin = 1,                ! Minimization (Yes)
   ntmin = 1,               ! Minimization Method (Steepest descent/Conjugate gradient)
   maxcyc = 5000,           ! Maximum number of minimization cycles (5000 cycles)
   ncyc = 1000,             ! Cycle of switch from steepest descent to conjugate gradient (at cycle
1000)
   cut = 10,                ! Non-bonding Cut-off (10 A)
   ntb = 1,                 ! Periodic Conditions (Yes)
 /
```

S06-Eql01-Heating-NTV.in

```
Restrained Heating 250 ps NVT MD
 &cntrl
   ig = -1,                 ! Pseudo-random number generator (random seed based on time)
   irest = 0,               ! Restart the Simulation? (No)
   ntx = 1,                 ! Read in only initial coordinates (ASCII)
   cut = 10,                ! Non-bonding Cut-off (10 A)
   ntc = 2,                 ! SHAKE bond length constraints (constrain bonds with H)
   ntf = 2,                 ! SHAKE force evaluation (omit bonds with H)
                            ! Note: SHAKE set for TIP-type waters (e.g. TIP3P)
   ntb = 1,                 ! PBC (Constant Volume)
   ntt = 3,                 ! Temperature scaling (Langevin dynamics)
   gamma_ln = 1.0,          ! Collision frequency (1 ps^-1)
   tempi = 0.0,             ! Initial temperature (O K, velocities assigned according to forces)
   temp0 = 100.0,           ! Reference temperature (100 K)
   ntr = 1,                 ! Harmonic constraints in Cartesian space (On)
   restraint_wt = 5.0,      ! Positional restraints weight ( 5 kcal/mol-A^2)
   restraintmask = ":1-213@CA,N,C,O",      ! Restrained atoms (protein backbone)
   dt = 0.002,              ! Simulation time-step (0.002 ps or 2 fs)
   nstlim = 25000,          ! Simulation length (25000 steps or 50 ps)
   ntpr = 1000,             ! Energy save interval (every 1000 steps or 2 ps)
   ntwx = 5000,             ! Coordinate/trajectory save interval (every 5000 steps or 10 ps)
   ntwr = 25000,            ! Restart file only at end of run.
```

```
  iwrap = 1,               ! Coordinates to be "wrapped" into primary box (on)
  ioutfm = 1,              ! Trajectory file format (Binary NetCDF)
  nmropt = 1,              ! Turn on NMR restraints - so we can control temp increase (see below).
 /
 &wt type = 'TEMP0',       ! Variable Conditions Type (Temp)
  istep1 = 0,              ! Start Change Step (0)
  istep2 = 25000,          ! Last Change Step (25000 steps or 50 ps)
  imult = 0                ! Interplation (Linear (Default))
  value1 = 0.0,            ! Start State (0 K)
  value2 = 100.0 /         ! End State (100 K)
 &wt type='END' /
```

S07-Eql02-Heating-NTP.in

```
Restrained Heating 250 ps NVT MD
 &cntrl
  ig = -1,                 ! Pseudo-random number generator (random seed based on time)
  irest = 1,               ! Restart the Simulation? (Yes)
  ntx = 5,                 ! Read coordinates, velocities, and box
  cut = 10,                ! Non-bonding Cut-off (10 A)
  ntc = 2,                 ! SHAKE bond length constraints (constrain bonds with H)
  ntf = 2,                 ! SHAKE force evaluation (omit bonds with H)
                           ! Note: SHAKE set for TIP-type waters (e.g. TIP3P)
  ntb = 2,                 ! PBC (Constant Pressure)
  ntp = 1,                 ! Constant Pressure MD (Isotropic position scaling)
  barostat = 1,            ! Berendsen Barostat used for equilibration
  pres0 = 1.0,             ! Reference Pressure (1 bar)
  taup = 5.0,              ! Pressure relaxation time (5 ps)
  ntt = 3,                 ! Temperature scaling (Langevin dynamics)
  gamma_ln = 1.0,          ! Collision frequency (1 ps^-1)
  tempi = 100.0,           ! Initial temperature
  temp0 = 300.0,           ! Reference temperature (300 K)
  ntr = 1,                 ! Harmonic constraints in Cartesian space (On)
  restraint_wt = 5.0,      ! Positional restraints weight ( 5 kcal/mol-A^2)
  restraintmask = ":1-213@CA,N,C,O",     ! Restrained atoms (protein backbone)
  dt = 0.002,              ! Simulation time-step (0.002 ps or 2 fs)
  nstlim = 100000,         ! Simulation length (100000 steps or 200 ps)
  ntpr = 1000,             ! Energy save interval (every 1000 steps or 2 ps)
  ntwx = 5000,             ! Coordinate/trajectory save interval (every 5000 steps or 10 ps)
  ntwr = 100000,           ! Restart file only at end of run.
  iwrap = 1,               ! Coordinates to be "wrapped" into primary box (on)
  ioutfm = 1,              ! Trajectory file format (Binary NetCDF)
  nmropt = 1,              ! Turn on NMR restraints - so we can control temp increase (see below).
 /
 &wt type = 'TEMP0',       ! Variable Conditions Type (Temp)
  istep1 = 0,              ! Start Change Step (0)
  istep2 = 75000,          ! Last Change Step (75000 steps or 150 ps)
  imult = 0                ! Interplation (Linear (Default))
  value1 = 100.0,          ! Start State (100 K)
  value2 = 300.0 /         ! End State (300 K)
 &wt type='END' /
```

S08-Eql03-EqlOnlyStage01.in

```
Restrained Equilibration Stage 1 250 ps NPT MD
 &cntrl
   ig = -1,                ! Pseudo-random number generator (random seed based on time)
   irest = 1,              ! Restart the Simulation? (Yes)
   ntx = 5,                ! Read coordinates, velocities, and box
   cut = 10,               ! Non-bonding Cut-off (10 A)
   ntc = 2,                ! SHAKE bond length constraints (constrain bonds with H)
   ntf = 2,                ! SHAKE force evaluation (omit bonds with H)
                           ! Note: SHAKE set for TIP-type waters (e.g. TIP3P)
   ntb = 2,                ! PBC (Constant Pressure)
   ntp = 1,                ! Constant Pressure MD (Isotropic position scaling)
   ntp = 1,                ! Constant Pressure MD (Isotropic position scaling)
   barostat = 1,           ! Berendsen Barostat used for equilibration
   pres0 = 1.0,            ! Reference Pressure (1 bar)
   taup = 5.0,             ! Pressure relaxation time (2 ps)
   ntt = 3,                ! Temperature scaling (Langevin thermostat) - Gives real dynamics
   gamma_ln = 5.0,         ! Collision frequency (5 ps^-1)
   temp0 = 300.0,          ! Reference temperature (300 K)
   ntr = 1,                ! Harmonic constraints in Cartesian space (On)
   restraint_wt = 5.0,     ! Positional restraints weight ( 5 kcal/mol-A^2)
   restraintmask = ":1-213@CA,N,C,O",     ! Restrained atoms (protein backbone)
   dt = 0.002,             ! Simulation time-step (0.002 ps or 2 fs)
   nstlim = 125000,        ! Simulation length (125000 steps or 250 ps)
   ntpr = 1000,            ! Energy save interval (every 1000 steps of 2 ps)
   ntwx = 5000,            ! Coordinate/trajectory save interval (every 5000 steps of 10 ps)
   ntwr = 125000,          ! Restart file only at end of run.
   iwrap = 1,              ! Coordinates to be "wrapped" into primary box (on)
   ioutfm = 1,             ! Trajectory file format (Binary NetCDF)
 /
```

S09-Eql04-EqlOnlyStage02.in

```
Unrestrained Equilibration Stage 2 500 ps NPT MD
 &cntrl
   ig = -1,                ! Pseudo-random number generator (random seed based on time)
   irest = 1,              ! Restart the Simulation? (Yes)
   ntx = 5,                ! Read coordinates, velocities, and box
   cut = 10,               ! Non-bonding Cut-off (10 A)
   ntc = 2,                ! SHAKE bond length constraints (constrain bonds with H)
   ntf = 2,                ! SHAKE force evaluation (omit bonds with H)
                           ! Note: SHAKE set for TIP-type waters (e.g. TIP3P)
   ntb = 2,                ! PBC (Constant Pressure)
   ntp = 1,                ! Constant Pressure MD (Isotropic position scaling)
   barostat = 2,           ! Monte Carlo Barostat - Optimal for GPU runs
   mcbarint = 1000,        ! Steps between volume changes for the barostat
   pres0 = 1.0,            ! Reference Pressure (1 bar)
   taup = 2.0,             ! Pressure relaxation time (2 ps)
   ntt = 3,                ! Temperature scaling (Langevin thermostat) - Gives real dynamics
   gamma_ln = 5.0,         ! Collision frequency (5 ps^-1)
   temp0 = 300.0,          ! Reference temperature (300 K)
```

```
   dt = 0.002,             ! Simulation time-step (0.002 ps or 2 fs)
   nstlim = 250000,        ! Simulation length (250000 steps or 250 ps)
   ntpr = 5000,            ! Energy save interval (every 5000 steps of 10 ps)
   ntwx = 5000,            ! Coordinate/trajectory save interval (every 5000 steps of 10 ps)
   ntwr = 250000,          ! Restart file only at end of run.
   iwrap = 1,              ! Coordinates to be "wrapped" into primary box (on)
   ioutfm = 1,             ! Trajectory file format (Binary NetCDF)
 /
```

S10-Pro01-MD_10ns.in

```
10 ns NTP MD
 &cntrl
   ig = -1,                ! Pseudo-random number generator (random seed based on time)
   irest = 1,              ! Restart the Simulation? (Yes)
   ntx = 5,                ! Read coordinates, velocities (ASCII)
   cut = 10,               ! Non-bonding Cut-off (10 A)
   ntc = 2,                ! SHAKE bond length constraints (constrain bonds with H)
   ntf = 2,                ! SHAKE force evaluation (omit bonds with H)
                           ! Note: SHAKE set for TIP-type waters (e.g. TIP3P)
   ntb=2,                  ! PBC (Constant Pressure)
   ntp = 1,                ! Constant Pressure MD (Isotropic position scaling)
   barostat = 2,           ! Monte Carlo Barostat - Optimal for GPU runs
   mcbarint = 1000,        ! Steps between volume changes for the barostat
   pres0 = 1.0,            ! Reference Pressure (1 bar)
   taup = 2.0,             ! Pressure relaxation time (2 ps)
   ntt = 3,                ! Temperature scaling (Langevin thermostat) - Gives real dynamics
   gamma_ln = 5.0,         ! Collision frequency (5 ps^-1)
   temp0 = 300.0,          ! Reference temperature (300 K)
   dt = 0.002,             ! Simulation time-step (0.002 ps or 2 fs)
   nstlim = 5000000,       ! Simulation length (5000000 steps or 10 ns)
   ntpr = 5000,            ! Energy save interval (every 5000 steps or 10 ps)
   ntwx = 5000,            ! Coordinate/trajectory save interval (every 5000 steps or 10 ps)
   ntwr = 5000,            ! Restart file save interval (every 5000 steps or 10 ps)
   iwrap = 1,              ! Coordinates to be "wrapped" into primary box (on)
   ioutfm = 1,             ! Trajectory file format (Binary NetCDF)
 /
```

S10-Pro01-MD_50ns.in

```
50 ns NTP MD
 &cntrl
   ig = -1,                ! Pseudo-random number generator (random seed based on time)
   irest = 1,              ! Restart the Simulation? (Yes)
   ntx = 5,                ! Read coordinates, velocities (ASCII)
   cut = 10,               ! Non-bonding Cut-off (10 A)
   ntc = 2,                ! SHAKE bond length constraints (constrain bonds with H)
   ntf = 2,                ! SHAKE force evaluation (omit bonds with H)
                           ! Note: SHAKE set for TIP-type waters (e.g. TIP3P)
   ntb=2,                  ! PBC (Constant Pressure)
   ntp = 1,                ! Constant Pressure MD (Isotropic position scaling)
   barostat = 2,           ! Monte Carlo Barostat - Optimal for GPU runs
```

```
   mcbarint = 1000,         ! Steps between volume changes for the barostat
   pres0 = 1.0,             ! Reference Pressure (1 bar)
   taup = 2.0,              ! Pressure relaxation time (2 ps)
   ntt = 3,                 ! Temperature scaling (Langevin thermostat) - Gives real dynamics
   gamma_ln = 5.0,          ! Collision frequency (5 ps^-1)
   temp0 = 300.0,           ! Reference temperature (300 K)
   dt = 0.002,              ! Simulation time-step (0.002 ps or 2 fs)
   nstlim = 25000000,       ! Simulation length (25000000 steps or 50 ns)
   ntpr = 5000,             ! Energy save interval (every 5000 steps or 10 ps)
   ntwx = 5000,             ! Coordinate/trajectory save interval (every 5000 steps or 10 ps)
   ntwr = 5000,             ! Restart file save interval (every 5000 steps or 10 ps)
   iwrap = 1,               ! Coordinates to be "wrapped" into primary box (on)
   ioutfm = 1,              ! Trajectory file format (Binary NetCDF)
 /
```

Scheme S2: Scripts used for MD simulation

References

(1) Brooijmans, N.; Kuntz, I. D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32* (1), 335–373.

(2) Antunes, D. A.; Devaurs, D.; Kavraki, L. E. Understanding the Challenges of Protein Flexibility in Drug Design. *Expert Opin. Drug Discov.* **2015**, *10* (12), 1301–1313.

(3) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260.

(4) Betz, R. M.; Walker, R. C. Paramfit: Automated Optimization of Force Field Parameters for Molecular Dynamics Simulations. *J. Comput. Chem.* **2015**, *36* (2), 79–87.

(5) Vanommeslaeghe, K.; Yang, M.; MacKerell, A. D., Jr. Robustness in the Fitting of Molecular Mechanics Parameters. *J. Comput. Chem.* **2015**, *36* (14), 1083.

(6) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296.

(7) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, NA – NA.

(8) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7* (12), 4026–4037.

(9) Ebejer, J.-P.; Morris, G. M.; Deane, C. M. Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* **2012**, *52* (5), 1146–1158.

(10) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572–584.

(11) Lexa, K. W.; Carlson, H. A. Protein Flexibility in Docking and Surface Mapping. *Q. Rev. Biophys.* **2012**, *45* (3), 301–343.

(12) Lin, J.-H. Accommodating Protein Flexibility for Structure-Based Drug Design. *Curr. Top. Med. Chem.* **2011**, *11* (2), 171–178.

(13) Spyrakis, F.; BidonChanal, A.; Barril, X.; Luque, F. J. Protein Flexibility and Ligand Recognition: Challenges for Molecular Modeling. *Curr. Top. Med. Chem.* **2011**, *11* (2), 192–210.
(14) Ellingson, S. R.; Miao, Y.; Baudry, J.; Smith, J. C. Multi-Conformer Ensemble Docking to Difficult Protein Targets. *J. Phys. Chem. B* **2015**, *119* (3), 1026–1034.
(15) Sørensen, J.; Demir, Ö.; Swift, R. V.; Feher, V. A.; Amaro, R. E. Molecular Docking to Flexible Targets. In *Methods in Molecular Biology*; 2014; pp 445–469.
(16) Huang, S.-Y.; Zou, X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins* **2007**, *66* (2), 399–421.
(17) Totrov, M.; Abagyan, R. Flexible Ligand Docking to Multiple Receptor Conformations: A Practical Alternative. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 178–184.
(18) Sørensen, J.; Demir, Ö.; Swift, R. V.; Feher, V. A.; Amaro, R. E. Molecular Docking to Flexible Targets. In *Methods in Molecular Biology*; 2014; pp 445–469.
(19) Wong, C. F. Flexible Receptor Docking for Drug Discovery. *Expert Opin. Drug Discov.* **2015**, *10* (11), 1189–1200.
(20) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Exploring Protein Flexibility: Incorporating Structural Ensembles from Crystal Structures and Simulation into Virtual Screening Protocols. *J. Phys. Chem. B* **2012**, *116* (23), 6952–6959.
(21) Henrich, S.; Salo- Ahen, O. M. H.; Huang, B.; Rippmann, F. F.; Cruciani, G.; Wade, R. C. Computational Approaches to Identifying and Characterizing Protein Binding Sites for Ligand Design. *J. Mol. Recognit.* **2010**, *23* (2), 209–219.
(22) Kokh, D. B.; Czodrowski, P.; Rippmann, F.; Wade, R. C. Perturbation Approaches for Exploring Protein Binding Site Flexibility to Predict Transient Binding Pockets. *J. Chem. Theory Comput.* **2016**, *12* (8), 4100–4113.
(23) Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C. Protein Binding Pocket Dynamics. *Acc. Chem. Res.* **2016**, *49* (5), 809–815.
(24) Kokh, D. B.; Richter, S.; Henrich, S.; Czodrowski, P.; Rippmann, F.; Wade, R. C. TRAPP: A Tool for Analysis of Transient Binding Pockets in Proteins. *J. Chem. Inf. Model.* **2013**, *53* (5), 1235–1252.
(25) Kleywegt, G. J.; Jones, T. A. Detection, Delineation, Measurement and Display of Cavities in Macromolecular Structures. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50* (Pt 2), 178–185.
(26) Desaphy, J.; Azdimousa, K.; Kellenberger, E.; Rognan, D. Comparison and Druggability Prediction of Protein–Ligand Binding Sites from Pharmacophore-Annotated Cavity Shapes. *J. Chem. Inf. Model.* **2012**, *52* (8), 2287–2299.
(27) Paramo, T.; East, A.; Garzón, D.; Ulmschneider, M. B.; Bond, P. J. Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: Trj_cavity. *J. Chem. Theory Comput.* **2014**, *10* (5), 2151–2164.
(28) Craig, I. R.; Pfleger, C.; Gohlke, H.; Essex, J. W.; Spiegel, K. Pocket-Space Maps to Identify Novel Binding-Site Conformations in Proteins. *J. Chem. Inf. Model.* **2011**, *51* (10), 2666–2679.
(29) Durrant, J. D.; Votapka, L.; Sørensen, J.; Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **2014**, *10* (11), 5047–5056.
(30) Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design. *Protein Sci.* **1998**, *7* (9), 1884–1897.
(31) Saberi Fathi, S.; Fathi, S. S.; Tuszynski, J. A. A Simple Method for Finding a Protein's Ligand-Binding Pockets. *BMC Struct. Biol.* **2014**, *14* (1), 18.
(32) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformatics* **2009**, *10*, 168.
(33) Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tufféry, P. Fpocket: Online Tools for Protein Ensemble Pocket Detection and Tracking. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W582–W589.
(34) Craig, I. R.; Pfleger, C.; Gohlke, H.; Essex, J. W.; Spiegel, K. Pocket-Space Maps to Identify Novel Binding-Site Conformations in Proteins. *J. Chem. Inf. Model.* **2011**, *51* (10), 2666–2679.
(35) Laurent, B.; Chavent, M.; Cragnolini, T.; Dahl, A. C. E.; Pasquali, S.; Derreumaux, P.; Sansom, M. S. P.;

Baaden, M. Epock: Rapid Analysis of Protein Pocket Dynamics. *Bioinformatics* **2015**, *31* (9), 1478–1480.
(36) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. POVME: An Algorithm for Measuring Binding-Pocket Volumes. *J. Mol. Graph. Model.* **2011**, *29* (5), 773–776.
(37) Durrant, J. D.; McCammon, J. A. BINANA: A Novel Algorithm for Ligand-Binding Characterization. *J. Mol. Graph. Model.* **2011**, *29* (6), 888–893.
(38) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9* (3), 10–20.
(39) Millman, K. J.; Jarrod Millman, K.; Aivazis, M. Python for Scientists and Engineers. *Comput. Sci. Eng.* **2011**, *13* (2), 9–12.
(40) Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J. An Automated Approach for Clustering an Ensemble of NMR-Derived Protein Structures into Conformationally Related Subfamilies. *Protein Eng.* **1996**, *9* (11), 1063–1065.
(41) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38.
(42) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
(43) Obermann, W. M.; Sondermann, H.; Russo, A. A.; Pavletich, N. P.; Hartl, F. U. In Vivo Function of Hsp90 Is Dependent on ATP Binding and ATP Hydrolysis. *J. Cell Biol.* **1998**, *143* (4), 901–910.
(44) Wright, L.; Barril, X.; Dymock, B.; Sheridan, L.; Surgenor, A.; Beswick, M.; Drysdale, M.; Collier, A.; Massey, A.; Davies, N.; Fink, A.; Fromont, C.; Aherne, W.; Boxall, K.; Sharp, S.; Workman, P.; Hubbard, R. E. Structure-Activity Relationships in Purine-Based Inhibitor Binding to HSP90 Isoforms. *Chem. Biol.* **2004**, *11* (6), 775–785.
(45) Brough, P. A.; Aherne, W.; Barril, X.; Borgognoni, J.; Boxall, K.; Cansfield, J. E.; Cheung, K.-M. J.; Collins, I.; Davies, N. G. M.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Finch, H.; Fink, A.; Hayes, A.; Howes, R.; Hubbard, R. E.; James, K.; Jordan, A. M.; Lockie, A.; Martins, V.; Massey, A.; Matthews, T. P.; McDonald, E.; Northfield, C. J.; Pearl, L. H.; Prodromou, C.; Ray, S.; Raynaud, F. I.; Roughley, S. D.; Sharp, S. Y.; Surgenor, A.; Walmsley, D. L.; Webb, P.; Wood, M.; Workman, P.; Wright, L. 4,5-Diarylisoxazole Hsp90 Chaperone Inhibitors: Potential Therapeutic Agents for the Treatment of Cancer. *J. Med. Chem.* **2008**, *51* (2), 196–218.
(46) Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G. M.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C.; Fromont, C.; Hayes, A.; Hubbard, R. E.; Jordan, A. M.; Jensen, M. R.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F. I.; Robertson, A.; Roughley, S. D.; Schoepfer, J.; Simmonite, H.; Sharp, S. Y.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L. Combining Hit Identification Strategies: Fragment-Based and in Silico Approaches to Orally Active 2-aminothieno[2,3-D]pyrimidine Inhibitors of the Hsp90 Molecular Chaperone. *J. Med. Chem.* **2009**, *52* (15), 4794–4809.
(47) Murray, C. W.; Carr, M. G.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S.; Coyle, J. E.; Downham, R.; Figueroa, E.; Frederickson, M.; Graham, B.; McMenamin, R.; O'Brien, M. A.; Patel, S.; Phillips, T. R.; Williams, G.; Woodhead, A. J.; Woolford, A. J.-A. Fragment-Based Drug Discovery Applied to Hsp90. Discovery of Two Lead Series with High Ligand Efficiency. *J. Med. Chem.* **2010**, *53* (16), 5942–5955.
(48) Roughley, S. D.; Hubbard, R. E. How Well Can Fragments Explore Accessed Chemical Space? A Case Study from Heat Shock Protein 90. *J. Med. Chem.* **2011**, *54* (12), 3989–4005.
(49) Miura, T.; Fukami, T. A.; Hasegawa, K.; Ono, N.; Suda, A.; Shindo, H.; Yoon, D.-O.; Kim, S.-J.; Na, Y.-J.; Aoki, Y.; Shimma, N.; Tsukuda, T.; Shiratori, Y. Lead Generation of Heat Shock Protein 90 Inhibitors by a Combination of Fragment-Based Approach, Virtual Screening, and Structure-Based Drug Design. *Bioorg. Med. Chem. Lett.* **2011**, *21* (19), 5778–5783.
(50) Barta, T. E.; Veal, J. M.; Rice, J. W.; Partridge, J. M.; Fadden, R. P.; Ma, W.; Jenks, M.; Geng, L.; Hanson, G. J.; Huang, K. H.; Barabasz, A. F.; Foley, B. E.; Otto, J.; Hall, S. E. Discovery of Benzamide Tetrahydro-4H-Carbazol-4-Ones as Novel Small Molecule Inhibitors of Hsp90. *Bioorg. Med. Chem. Lett.* **2008**, *18* (12), 3517–3521.
(51) Cho-Schultz, S.; Patten, M. J.; Huang, B.; Elleraas, J.; Gajiwala, K. S.; Hickey, M. J.; Wang, J.; Mehta, P. P.;

Kang, P.; Gehring, M. R.; Kung, P.-P.; Sutton, S. C. Solution-Phase Parallel Synthesis of Hsp90 Inhibitors. *J. Comb. Chem.* **2009**, *11* (5), 860–874.
(52) Kung, P.-P.; Huang, B.; Zhang, G.; Zhou, J. Z.; Wang, J.; Digits, J. A.; Skaptason, J.; Yamazaki, S.; Neul, D.; Zientek, M.; Elleraas, J.; Mehta, P.; Yin, M.-J.; Hickey, M. J.; Gajiwala, K. S.; Rodgers, C.; Davies, J. F.; Gehring, M. R. Dihydroxyphenylisoindoline Amides as Orally Bioavailable Inhibitors of the Heat Shock Protein 90 (hsp90) Molecular Chaperone. *J. Med. Chem.* **2010**, *53* (1), 499–503.
(53) Zapf, C. W.; Bloom, J. D.; Li, Z.; Dushin, R. G.; Nittoli, T.; Otteng, M.; Nikitenko, A.; Golas, J. M.; Liu, H.; Lucas, J.; Boschelli, F.; Vogan, E.; Olland, A.; Johnson, M.; Levin, J. I. Discovery of a Stable Macrocyclic O-Aminobenzamide Hsp90 Inhibitor Which Significantly Decreases Tumor Volume in a Mouse Xenograft Model. *Bioorg. Med. Chem. Lett.* **2011**, *21* (15), 4602–4607.
(54) Kung, P.-P.; Sinnema, P.-J.; Richardson, P.; Hickey, M. J.; Gajiwala, K. S.; Wang, F.; Huang, B.; McClellan, G.; Wang, J.; Maegley, K.; Bergqvist, S.; Mehta, P. P.; Kania, R. Design Strategies to Target Crystallographic Waters Applied to the Hsp90 Molecular Chaperone. *Bioorg. Med. Chem. Lett.* **2011**, *21* (12), 3557–3562.
(55) Casale, E.; Amboldi, N.; Brasca, M. G.; Caronni, D.; Colombo, N.; Dalvit, C.; Felder, E. R.; Fogliatto, G.; Galvani, A.; Isacchi, A.; Polucci, P.; Riceputi, L.; Sola, F.; Visco, C.; Zuccotto, F.; Casuscelli, F. Fragment-Based Hit Discovery and Structure-Based Optimization of Aminotriazoloquinazolines as Novel Hsp90 Inhibitors. *Bioorg. Med. Chem.* **2014**, *22* (15), 4135–4150.
(56) Davies, N. G. M.; Browne, H.; Davis, B.; Drysdale, M. J.; Foloppe, N.; Geoffrey, S.; Gibbons, B.; Hart, T.; Hubbard, R.; Jensen, M. R.; Mansell, H.; Massey, A.; Matassova, N.; Moore, J. D.; Murray, J.; Pratt, R.; Ray, S.; Robertson, A.; Roughley, S. D.; Schoepfer, J.; Scriven, K.; Simmonite, H.; Stokes, S.; Surgenor, A.; Webb, P.; Wood, M.; Wright, L.; Brough, P. Targeting Conserved Water Molecules: Design of 4-Aryl-5-cyanopyrrolo[2,3-D]pyrimidine Hsp90 Inhibitors Using Fragment-Based Screening and Structure-Based Optimization. *Bioorg. Med. Chem.* **2012**, *20* (22), 6770–6789.
(57) Chen, D.; Shen, A.; Li, J.; Shi, F.; Chen, W.; Ren, J.; Liu, H.; Xu, Y.; Wang, X.; Yang, X.; Sun, Y.; Yang, M.; He, J.; Wang, Y.; Zhang, L.; Huang, M.; Geng, M.; Xiong, B.; Shen, J. Discovery of Potent N-(isoxazol-5-Yl)amides as HSP90 Inhibitors. *Eur. J. Med. Chem.* **2014**, *87*, 765–781.
(58) Ren, J.; Yang, M.; Liu, H.; Cao, D.; Chen, D.; Li, J.; Tang, L.; He, J.; Chen, Y.-L.; Geng, M.; Xiong, B.; Shen, J. Multi-Substituted 8-aminoimidazo[1,2-A]pyrazines by Groebke-Blackburn-Bienaymé Reaction and Their Hsp90 Inhibitory Activity. *Org. Biomol. Chem.* **2015**, *13* (5), 1531–1535.
(59) McBride, C. M.; Levine, B.; Xia, Y.; Bellamacina, C.; Machajewski, T.; Gao, Z.; Renhowe, P.; Antonios-McCrea, W.; Barsanti, P.; Brinner, K.; Costales, A.; Doughan, B.; Lin, X.; Louie, A.; McKenna, M.; Mendenhall, K.; Poon, D.; Rico, A.; Wang, M.; Williams, T. E.; Abrams, T.; Fong, S.; Hendrickson, T.; Lei, D.; Lin, J.; Menezes, D.; Pryer, N.; Taverna, P.; Xu, Y.; Zhou, Y.; Shafer, C. M. Design, Structure-Activity Relationship, and in Vivo Characterization of the Development Candidate NVP-HSP990. *J. Med. Chem.* **2014**, *57* (21), 9124–9129.
(60) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
(61) *Gaussian 09*; Gaussian, Inc. Wallingford CT., 2009.
(62) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39* (suppl), W511–W517.
(63) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821.
(64) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280.
(65) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725.

(66) Kendall, M. G. A NEW MEASURE OF RANK CORRELATION. *Biometrika* **1938**, *30* (1-2), 81–93.
(67) Vauquelin, G.; Charlton, S. J. Long-Lasting Target Binding and Rebinding as Mechanisms to Prolong in Vivo Drug Action. *Br. J. Pharmacol.* **2010**, *161* (3), 488–508.