MDO: A Computational Protocol for Accurate Prediction of Full Flexibility Receptor-Ligand Binding Mode Structures

Amar Y. Al-Ansi
University of Science and Technology of China

Zijing Lin (✉ zjlin@ustc.edu.cn)
University of Science and Technology of China

Research Article

Keywords: molecular dynamics, molecular docking, clustering analysis, QM/MM, structure based drug design

Posted Date: September 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-892370/v1

License: ☕️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Predicting the binding structure of bio-complex is essential for understanding its properties, functions, and mechanisms, but is rather difficult due to the huge sampling space involved. A new computational protocol, MDO, for finding the ligand binding structure is proposed. MDO consists of global sampling via MD simulation and clustering of the receptor configurations, local sampling via molecular docking and clustering of the ligand conformations, and binding structure optimization by the ONIOM (QM/QM) method. MDO is tested on 15 protein-ligand complexes with known accurate structures. The success rate of MDO predictions, with RMSD < 2 Å, is found to be 67%, substantially higher than the 40% success rate of conventional methods. The MDO success rate can be increased to 83% if the ONIOM calculations are applied only for the starting poses with ligands inside the binding cavities. The MDO protocol is a promising tool for the structure based drug design.

1. Introduction

The principal aim of rational drug design and development is to detect a small molecule that binds robustly to a particular bio-macro-molecule. However, the identification of bio-complex binding structures is still a prime challenge for the structure-based drug design (SBDD). The difficulty with the structural determinations is simply due to the complicated potential energy landscape of the bio-complex with an essentially infinite number of local minima.

The molecular docking technique is a simple way to obtain the likely binding mode of the bio-complex system and estimate its binding affinity. In this technique, a ligand is first bound into a receptor in different binding mode conformations, then the conformations are ranked with a scoring function. During the docking process, seeking the ligand conformations for a fixed receptor structure, i.e., rigid docking, is very common, while the exploration of receptor flexibility or flexible docking remains challenging and rare. However, the importance of receptor flexibility for SBDD by modifying its binding mode based on the ligand orientation is well documented and the flexible docking should be pursued if possible.

A diversity of flexible docking approaches have been suggested, including the ensemble docking with discrete receptor conformations, the induced-fit that allows the receptor conformation to change during docking, and the modification of receptor-ligand interaction to simplify the potential energy landscape of the bio-complex. While some flexible dockings focus on the flexible receptor side chain, others have limited backbone movements. As an example of special interest, the recently released AutoDock 4.0 software allows the side chains of macromolecule to be flexible. AutoDock 4.0 is also fast and capable of performing 40,000 rigid dockings in a day on one CPU. The docking algorithm is capable of finding diverse binding modes, but fails to find the correct binding mode through the scoring function. The success rate of docking ranges from 35–60% when judged by the Root Mean Square Deviation (RMSD) of less than 2 Å.
Generally, the success of molecular docking depends on the positive identification of binding modes and accurate estimation of their free energies. Unfortunately, most docking methods employ molecular mechanics (MM) force fields that are inadequate for the description of polarization, charge transfer, bond breaking/forming, etc. Consequently, recourse to use quantum mechanics (QM) based methods, especially the density functional theory (DFT), is desirable for the improvement of molecular docking.

However, the computational cost of QM is too high for the whole bio-complex system. Some compromise between the needs of numerical accuracy and computational efficiency is required, leading to the QM/MM technique. In the QM/MM approach, the ligand, and occasionally important atoms of the receptor, is treated by the QM method, while the rest of the receptor-ligand complex is treated by the MM method. In particular, ONIOM (our N-layered Integrated Molecular Orbital and Molecular Mechanics) is a highly successful QM/MM technique for large bio-complex systems. ONIOM uses different levels of theory for different layers of the bio-complex system and generates reliable system geometry and energy in a short time. ONIOM is used in computational protocols like DOX and CSAMP to find the optimal binding modes of rigid docking.

This paper introduces a new computational protocol, MDO, to obtain the binding mode of fully flexible receptor-ligand complex, i.e., both the side chain and backbone of the receptor are flexible. MDO consists of global sampling of protein conformations by molecular dynamics (MD) simulations, local sampling of the ligand conformation space by rigid docking, and binding mode prediction by geometry optimization of the ligand-protein binding poses using the ONIOM method. The advantage of the MDO protocol is demonstrated by comparing the theoretical binding mode structures of 15 crystal bio-complex systems with the experimental results.

2. Computational Method

2.1 The MDO protocol

Figure 1 is a schematic of the new computational Protocol, MDO, for the prediction of full flexibility receptor-ligand binding mode structures. As shown in Fig. 1, there are three major steps in MDO: 1) global sampling, 2) local sampling, 3) binding geometry optimization and determination. The details of the three steps are described below.

2.1.1 Global sampling: MD simulations of ligand-receptor system

MD simulations are performed on the interested ligand-receptor bio-complex to generate a large pool of receptor conformations. All the MD trajectories are used in geometric clustering analysis to generate the most representative ligand-protein conformations.

In this work, each bio-complex system is subjected to MD simulation for 10 ns, using Gromacs 2018 package, periodic boundary conditions (PBC), and CHARMM36 force field. Ligand topologies are
generated using SwissParam server\textsuperscript{34} that is compatible with the CHARMM force fields.

The bio-complex systems were solvated in a cubic box with a TIP3P\textsuperscript{35} model of water molecules as a solvent, with 1 Å between the bio-complex and the edge of the box. The solvated systems were naturalized by interchanging sodium or chloride ions with the solvent molecules. The systems were relaxed using the steepest descent method with maximum force convergence of 10 kJ mol\(^{-1}\) nm\(^{-1}\). The LINCS algorithm was applied to constrain all-bonds\textsuperscript{36}. The relaxed systems were heated gradually for 100 ps.

The MD runs use the leap-frog integration\textsuperscript{37}, specifically the leap-frog integrator\textsuperscript{38}, with the time step of 2 fs, to integrate the equations of motion. The temperature control uses a weak coupling of V-rescale algorithm with a time constant of 0.1 ps\textsuperscript{39}, pressure coupling used the Berendsen algorithm with a time constant of 0.2 ps\textsuperscript{39}. The pressure control uses in the Parrinello-Rahman algorithm\textsuperscript{40} with a time constant of 0.1 ps. Long-range interactions were treated using Particle Mesh Ewald (PME) method\textsuperscript{41,42}, with 4-order interpolation, grid spacing for FFT of 0.16 nm, and coulomb radius of 102 nm. The short-range of cut-off distances set to 1.2 nm for calculation of van der Waals (VDW) interactions\textsuperscript{43}. Randomly initial velocities were generated, using the NPT equilibration over 100 ps.

The MD frames in an interval of 10 ps are saved for analysis. The geometry clustering analysis of the MD trajectories is carried out using the Gromos method\textsuperscript{44}, after removing the PBC in the MD runs. The Gromos method uses the alpha carbon (C-α) RMSD to characterize the similarity of MD frames. The Gromos algorithm counts the number of neighbors of the frames using an RMSD threshold, discovers the frame with the largest number of neighbors as the frames centroid, removes the frames from the pool, and repeats the procedure for the remaining frames. While the top clusters represent the highest probabilities of the MD frames, only one frame for each cluster representing the cluster centroid is extracted as a PDB file for use in the molecular docking study. The RMSD threshold is varied between 0.1 nm and 0.15 nm to find the optimal threshold\textsuperscript{45} for each ligand-receptor bio-complex in this study.

2.1.2 Local sampling: rigid docking and clustering

Molecular docking is carried out for each clustering structure obtained from the MD simulations to generate more suitable ligand conformations. Information about the presence of atoms in the chosen receptor conformation is used to avoid steric clashes and bad geometry of the ligand. Ligand conformations obtained during docking simulation were clustered using the Ensemble Cluster program. This program clusters the members of the ligand conformational ensemble and identifies the cluster representatives using the pair-wise best fit RMSDs.

The molecular docking simulations are performed with the DOCK program in the Molecular Operating Environment (MOE)\textsuperscript{46}. Ligand conformations are generated with the bond rotation method. The docking is accomplished by using the DOCK program with the default setting, i.e. ‘Triangle matcher/London dG’ and ‘Rigid Receptor/forcefield/GBVI-WAS dG’ parameters were chosen as placement and refinement
methods, respectively. The docking generates ~ 200 ligand conformations for every representative of the 10 top MD frames, i.e. a total of ~ 2000 ligand conformations are generated for every receptor.

Clustering of ligand conformations obtained from the docking simulations is performed using the ensemble cluster tool in the UCSF-Chimera software. This tool applies the approach of Kelley et al. The ensemble of conformations is grouped into clusters of similar conformations. One conformation from each cluster is selected to exemplify the conformations in that cluster. After clustering, the top 10 binding poses for each of the 10 most populated protein conformations, or a total of 100 binding poses, are left to represent the likely ligand-receptor binding modes.

2.1.3 Geometry optimization and ranking of ligand-protein binding poses

The top 100 most populated binding modes obtained above for each ligand-protein complex are subjected to geometry optimization using the ONIOM method. The binding mode structures are then ranked by the final ONIOM energy. The top ranked binding structure is taken as the best ligand-receptor binding mode structure.

The ONIOM calculations are carried out at the level of wB97X-D/6-31G*: PM6. All amino acid (AA) residues within 6 A of ligand atoms are included in the ONIOM calculation. There are about 700 atoms in the AA residues of the ONIOM system in this study. The ONIOM system is divided into two layers, a high-level part for the ligand, treated at the wB97X-D/6-31G* level, and a low-level part for the selected AA residues, treated at the PM6 level. The geometry optimizations of the large bio-systems are performed with the "loose" convergence criterion. The ONIOM calculations are conducted with the Gaussian09 suite of software.

2.2 Testing set of bio-complex systems

The proposed computational protocol MDO is tested on fifteen protein-ligand complexes whose x-rays crystallographic structures are available in the protein data bank (PDB). The PDB IDs of the 15 bio-complexes are: 1ERB, 1FKI, 1HW9, 2IFB, 1FKG, 1HW8, 3G15, 1HW1, 21OA, 3MDX, 2Q13, 1HWJ, 1LF2, 1HWK, 1LF3. When there are more than one ligands in a bio-complex, the first ligand is chosen for the study. Basic information about the 15 ligands can be found in Table S1 of the Supporting Information. In addition to having accurate 3D crystallographic structures in PDB, the 15 bio-complex systems are bound by non-covalent interactions. When preparing the bio-complexes for testing the MDO protocol, water molecules and ions presented in the crystal structures are detached. Missing atoms or residues that should exist in the crystallographic structure are added. Each bio-complex is assigned at the physiological conditions for titratable residues using preparation and 3D-protonation module of MOE.

Considering the flexible nature of MDO and for more comparison, we also perform flexible (induced fit) docking with the DOCK software on the bio-complexes, using Charmm27 for the energy minimization. Like the rigid docking, the bond rotation method is also used in the induced fit for the generation of ligand
conformations. The induced fit uses the settings 'Triangle matcher/London dG' and 'Induced-Fit/forcefield/GBVI-WAS dG' for the conformation placement and refinement, respectively.

3. Results And Discussion

3.1 MD simulation and clustering

Figure 2 shows the variations of RSMDs with the MD simulations for the 15 bio-complex systems. The PDB crystallographic structures are used as the references for the RMSD evaluations. The RMSDs for a bio-complex reflect implicitly the movements of the backbone, side chains and ligand of the bio-complex. As seen in Fig. 3, all RMSDs level off after ~ 2 ns of MD runs. That is, after relaxation from their crystal x-ray structures, the 15 bio-complex systems are quite stable against the MD runs under the control conditions. The backbone and C-α RMSDs for most proteins are under 2 Å, indicating high similarities between the MD frames and the original crystallographic structures.

The MD simulations yield a large number of protein configuration snapshots. The MD trajectories are subjected to clustering analysis with the C-α RMSD cut-off of 0.13, 0.11, 0.13, 0.12, 0.11, 0.13, 0.13, 0.12, 0.13, 0.13, 0.13, 0.13, 0.13, and 0.14 nm for 1ERB, 1FKI, 1HW9, 2IFB, 1FKG, 1HW8, 3GI5, 1HWI, 2I0A, 3MXD, 2QI3, 1HWJ, 1LF2, 1HWK, and 1LF3, respectively. The RMSD cut-offs are chosen to ensure the diversity of clusters, i.e., as many clusters as possible, while requiring the top 10 clusters to include ~ 90% or more of the MD trajectories. With the above used C-α RMSD cut-offs, the top 10 clusters are highly representative as they account for over 90% of the total MD frames for each of the 15 bio-complexes, as shown in Fig. 3. Details about the dependences of the number of clusters and the ratio of members in the top ten clusters on the C-α RMSD cut-off for each of the 15 bio-complexes can be found in Table S2 of the Supporting Information.

3.2 Binding mode prediction

Table 1 compares the binding mode RMSDs of different prediction methods. The RMSD score is calculated between the crystal binding pose as the reference and the predicted top scored pose. As seen in Table 1, only 6 out of the 15 top scored binding poses of DOCK rigid docking have RMSDs below 2 Å. The success rate of DOCK rigid docking is quite low at 40%. The performance of the Induced-Fit method is similar, also with 6 of the 15 RMSDs below 2 Å. However, the average RMSD of the Induced-Fit, 2.46 Å, is larger than the rigid docking, 2.10 Å. The smallest and largest RMSDs of the rigid docking are 1.15 Å for 2IFB and 3.49 Å for 1LF3, respectively. In comparison, the smallest and largest RMSDs of the Induced-Fit are respectively 1.20 Å for 2IFB and 5.18 Å for 1LF2.

The quality of the top binding poses is much improved with the MDO prediction. The success rate of MDO is 67%, with 10 of the 15 RMSDs below 2 Å. Comparison of the starting and resulting RMSDs of the MDO binding poses in Table 1 illustrates the power and the critical importance of the ONIOM calculations in determining the best binding poses. Here, the MDO starting pose refers to the starting structure from which the MDO predicted top pose was generated, i.e., the structure of the MDO pose before the ONIOM
calculation. As shown in Table 1, only 6 of the 15 MDO starting RMSDs are less than 2 Å. The overall quality of the MDO starting poses, with an average RMSD of 2.54 Å, is the worst among the methods shown in Table 1, worse than that of the DOCK rigid docking and the Induced-Fit. However, the geometry optimization by ONIOM reduces the RMSD for 12 of the 15 MDO starting poses, leading to a rate of 80% for the improvement of the binding structures. The MDO predictions are overall far superior to their starting structures. They are also clearly better than the predictions of the rigid docking and Induced-Fit.

Table 1
Comparison of RMSDs (in Å) of top binding mode structures of 15 Bio-complex systems as predicted by DOCK (Rigid, Induced-Fit) and MDO (Rigid, flexible).

| PDB ID | Rigid RMSD | Induced-Fit RMSD | MDO Binding Mode | starting RMSD | MDO RMSD |
|--------|------------|------------------|------------------|---------------|----------|
| 1ERB   | 1.27       | 1.37             | 7–10             | 1.36          | 0.74     |
| 1FKI   | 2.00       | 1.40             | 6 – 2            | 1.69          | 0.82     |
| 1HW9   | 2.38       | 2.70             | 10 – 6           | 2.83          | 1.09     |
| 2IFB   | 1.15       | 1.20             | 4 – 1            | 1.69          | 1.21     |
| 1FKG   | 2.48       | 1.47             | 7 – 4            | 3.15          | 1.36     |
| 1HW8   | 1.62       | 1.64             | 7 – 4            | 3.11          | 1.41     |
| 3GI5   | 2.70       | 2.13             | 6–7              | 2.94          | 1.54     |
| 1HWI   | 2.43       | 2.45             | 4 – 1            | 1.71          | 1.55     |
| 2I0A   | 2.29       | 2.99             | 10 – 2           | 2.24          | 1.77     |
| 3MXD   | 1.72       | 4.42             | 5 – 1            | 3.29          | 1.82     |
| 2QI3   | 2.02       | 2.28             | 3 – 2            | 2.97          | 2.41     |
| 1HWJ   | 1.59       | 1.36             | 10 – 4           | 3.68          | 2.65     |
| 1LF2   | 2.89       | 5.18             | 5 – 2            | 1.78          | 3.33     |
| 1HWK   | 1.42       | 2.97             | 5 – 3            | 1.61          | 3.93     |
| 1LF3   | 3.49       | 3.38             | 5–8              | 4.02          | 6.98     |

The MDO starting poses are usually ranked low by the scoring function of DOCK. 7 of the 15 top MDO starting poses are outside of the top 100 ranked rigid docking poses, as shown in Tables S7, S8, S11 - S15. The need to substantially improve the DOCK scoring function is evident. While the MDO prediction does not necessarily pick the best pose, it is usually among the best ones. In particular, for the 10 successful predictions, namely with RMSD < 2 Å, 7 of the MDO picks are the very best, 2 of them are the second best, and 1 of them is the third best, see Tables S3 – S12. That is, most of the MDO predictions
corresponds to the best choices, and all of them are among the top three structures. Moreover, the quality of MDO structure, when it is not the best pick, is not far from the best one. For example, the RMSDs of the MDO and best poses are respectively 0.74 Å and 0.71 Å for the ligand ETR, 0.82 Å and 0.68 Å for the ligand SB1, 1.55 Å and 1.36 Å for the ligand 115. The results are quite comforting and lend confidence to the MDO picks.

While the overall success of MDO is evident, the three cases of ILF2, 1HWK and 1LF3 shown in Table 1 are quite troubling. The MDO binding poses for the three bio-complexes are worse than their starting poses. To illustrate why the ONIOM calculations fail for the three cases but succeed in all the other cases, the starting MDO poses of 6 representative bio-systems are shown in Fig. 4. As seen in Fig. 4, the MDO starting poses for 1ERB, 2QI3 and 1HWJ are inside the binding cavities of the proteins. That is also the case for all the other studied bio-complexes, except for ILF2, 1HWK and 1LF3. The binding modes with the starting ligand poses inside the protein binding cavities are all improved by the ONIOM calculations. In the cases of 1LF2, 1HWK and 1LF3, as shown in Fig. 4(d, e, f), the MDO starting poses selected by the ensemble clustering of rigid docking poses of DOCK are out of the protein binding cavities. That is, the ensemble cluster fails to select the right poses. Due to the lack of proper ligand-receptor interactions, the ONIOM geometry optimizations lead to higher RMSDs than the starting ones.

The observations derived from Fig. 4 can be used to improve the success rate of MDO and/or avoid unnecessary ONIOM calculations. The MDO starting poses should be inspected before the ONIOM calculations to see if the ligands are inside the binding cavities. If positive, the starting poses are subjected to geometry optimizations by ONIOM. For the bio-complexes examined in this study, the rate of structural improvement is then increased to 100%, and the rate of successful binding pose determinations is increased to 83%. If negative, the starting poses can be skipped to avoid the meaningless ONIOM calculations.

For a bio-complex with poor binding poses selected by the ensemble clustering, the worst case scenario is to say that the MDO methodology is inapplicable to the bio-complex. However, additional efforts may be made to enable the use of MDO for this bio-complex. For example, the step of rigid docking by DOCK, and even the step of MD simulations, may be repeated to yield the suitable MDO starting poses. Once such starting poses are found, the chance of successful identification of binding pose is quite high, as inferred from the results for the 12 bio-complexes.

4. Conclusions

A new computational protocol, MDO, for the prediction of ligand binding structure is proposed. MDO consists of MD simulations for sampling the receptor configurations and clustering, rigid docking for sampling the ligand conformations and clustering, and ranking of the binding modes by the ONIOM method. MDO is tested on 15 protein-ligand complexes with known accurate structures. The MDO protocol yields RMSD < 2 Å for 10 of the 15 tested cases, or a success rate of 67%. In comparison, the success rates of both the rigid docking and induced-fit with DOCK are only 40% for the same 15 bio-
complexes. A key element for the success of MDO is the binding structure improvement by the ONIOM calculations. For the starting poses with the ligands inside the binding cavities, the MDO predictions are successful for 10 of the 12 bio-complexes, or a success rate of 83%, while the ONIOM structural improvement is observed for all the 12 bio-complexes. Therefore, the applicability of MDO can be predetermined by inspecting the binding poses obtained by rigid docking. The success rate of MDO predictions for applicable bio-complexes is over 80%, while the rate of structural improvement by the ONIOM calculations is 100%. MDO is a robust technique to identify the binding mode structures that are essential for the deep understanding of properties, functions, and mechanisms of bio-complex systems. MDO is a promising tool for the structure based drug design.

Declarations

ACKNOWLEDGEMENTS

The financial support of the National Natural Science Foundation of China (11774324, 12074362 & 11574284) and the computing time of the Supercomputing Center of the University of Science and Technology of China are gratefully acknowledged.

References

1. Jensen, F. Introduction to Computational Chemistry; Wiley, 2017, 3rd edit, Chapter 13, pp 404-444.
2. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Nat Rev Drug Discov 2004, 3(11), 935-949.
3. Reddy, A. S.; Pati, S. P.; Potukuchi, P. K.; Pradeep, H. N.; Sastry, G. N. Curr Protein Pept Sc 2007, 8(4), 329-351.
4. Ryde, U.; Soderhjelm, P. Chem Rev 2016, 116(9), 5520-5566.
5. Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y. B.; Humblet, C. J Chem Inf Model 2009, 49(6), 1455-1474.
6. Pagadala, N. S.; Syed, K.; Tuszynski, J. Biophys Rev 2017, 9(2), 91-102.
7. Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. J Med Chem 2006, 49(20), 5912-5931.
8. Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. Curr Opin Struc Biol 2002, 12(2), 197-203.
9. Barnard, D.; Diaz, B.; Hettich, L.; Chuang, E.; Zhang, X. F.; Avruch, J.; Marshall, M. Oncogene 1995, 10(7), 1283-1290.
10. Lexa, K. W.; Carlson, H. A. Q Rev Biophys 2012, 45(3), 301-343.
11. Yuriev, E.; Holien, J.; Ramsland, P. A. J Mol Recognit 2015, 28(10), 581-604.
12. Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. J Am Chem Soc 2002, 124(20), 5632-5633.
13. Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Proteins-Structure Function and Genetics 2002, 46(1), 34-40.
14. Corbeil, C. R.; Englebienne, P.; Moitessier, N. J Chem Inf Model 2007, 47(2), 435-449.
15. Huang, S. Y.; Zou, X. Q. Int J Mol Sci 2010, 11(8), 3016-3034.
16. Mizutani, M. Y.; Takamatsu, Y.; Ichinose, T.; Nakamura, K.; Itai, A. Proteins 2006, 63(4), 878-891.
17. Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; L'Heureux, L.; Moitessier, N. J Chem Inf Model 2008, 48(4), 902-909.
18. Zhao, Y.; Sanner, M. F. Proteins 2007, 68(3), 726-737.
19. Zhao, Y.; Sanner, M. F. J Comput Aid Mol Des 2008, 22(9), 673-679.
20. AutoDock. What's new?; http://autodock.scripps.edu/ (accessed October 26, 2020).
21. Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalska, K. J Comput Chem 2011, 32(4), 742-755.
22. Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M. Drug Discov Today 2007, 12(17-18), 725-731.
23. Zhou, T.; Huang, D. Z.; Caflisch, A. Curr Top Med Chem 2010, 10(1), 33-45.
24. Lin, H.; Truhlar, D. G. Theor Chem Acc 2007, 117(2), 185-199.
25. Senn, H. M.; Thiel, W. Angew Chem Int Ed 2009, 48(7), 1198-1229.
26. Rao, L.; Chi, B.; Ren, Y. L.; Li, Y. J.; Xu, X.; Wan, J. J Comput Chem 2016, 37(3), 336-344.
27. Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z. F.; Liu, F. Y.; Li, H. B.; Ding, L. N.; Morokuma, K. Chem Rev 2015, 115(12), 5678-5796.
28. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. J Phys Chem-U 1996, 100(50), 19357-19363.
29. Morokuma, K. Philos T Roy Soc A 2002, 360(1795), 1149-1164.
30. Wei, L.; Chi, B.; Ren, Y. L.; Rao, L.; Wu, J.; Shang, H.; Liu, J. Q.; Xiao, Y. T.; Ma, M. H.; Xu, X.; Wan, J. J Chem Theory Comput 2019, 15(7), 4264-4279.
31. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. J Comput Chem 2005, 26(16), 1701-1718.
32. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. SoftwareX 2015, 1-2, 19-25.
33. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. J Chem Theory Comput 2012, 8(9), 3257-3273.
34. Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. J Comput Chem 2011, 32(11), 2359-2368.
35. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phys 1983, 79(2), 926-935.
36. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J Comput Chem 1997, 18(12), 1463-1472.
37. Van Gunsteren, W. F.; Berendsen, H. J. C. Mol Simulat 1988, 1(3), 173-185.
38. Verlet, L. Phys Rev 1967, 159(1), 98-103.
39. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. J Chem Phys 1984, 81(8), 3684-3690.
40. Parrinello, M.; Rahman, A. J Appl Phys 1981, 52(12), 7182-7190.
41. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98(12), 10089-10092.
42. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. J Chem Phys 1995, 103(19), 8577-8593.
43. M.J Abraham, D. v. d. S., E. Lindahl, B. Hess, and the GROMACS development team, GROMACS User Manual version, 2018, Chapter 4, pp 110-112.
44. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Angew Chem Int Edit 1999, 38(1-2), 236-240.
45. Peng, L. X., The role of computer simulations in the preclinical development of semiflexible polymeric anticancer therapeutics. Ph.D. Thesis, University of California, 2010.
46. Molecular Operating Environment (MOE), 2015.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2015.
47. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. J Comput Chem 2004, 25(13), 1605-1612.
48. Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J. Protein Eng 1996, 9(11), 1063-1065.
49. Chai, J. D.; Head-Gordon, M. Phys Chem Chem Phys 2008, 10(44), 6615-6620.
50. Stewart, J. J. P. J Mol Model 2007, 13(12), 1173-1213.
51. Foresman, J. B.; Frisch, A. E.; Gaussian, I. Exploring Chemistry with Electronic Structure Methods; Gaussian, 1996, Chapter 3, pp 85-138.
52. Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2013.
53. RCSB PDB http://www.rcsb.org/ (accessed October 26, 2020).
54. Zanotti, G.; Malpeli, G.; Berni, R. J Biol Chem 1993, 268(33), 24873-24879.
55. Holt, D. A.; Luengo, J. I.; Yamashita, D. S.; Oh, H. J.; Konialian, A. L.; Yen, H. K.; Rozamus, L. W.; Brandt, M.; Bossard, M. J.; Levy, M. A.; Eggleston, D. S.; Liang, J.; Schultz, L. W.; Stout, T. J.; Clardy, J. J Am Chem Soc 1993, 115(22), 9925-9938.
Figures
Figure 1

Schematic of the MDO Protocol.
Figure 2

RMSDs of the protein backbone, C-alpha, side chain, and the ligand during MD simulations.
Figure 3

C-α RMSD cut-off of protein MD frames, total number of clusters, and percentage of the MD frames of top ten clusters.
Figure 4
Binding modes of bio-complexes: a) 1ERB, b) 2QI3, c) 1HWJ, d) 1LF2, e) 1HWK, f) 1LF3. The ligand binding poses obtained by MD simulations are shown in red, while the MDO starting poses are in yellow.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- GraphicalAbstract.jpg
- supportinginformation.doc