

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Fragment Dissolved Molecular Dynamics: A systematic and efficient method to locate binding sites

Journal:	<i>Journal of Chemical Information and Modeling</i>
Manuscript ID	ci-2020-00770d
Manuscript Type:	Article
Date Submitted by the Author:	07-Jul-2020
Complete List of Authors:	Privat, Cristian; Universitat de Barcelona, Ciència dels Materials i Química Física Granadino-Roldan, Jose; Universidad de Jaen, Química Física y Analítica Bonet, Jordi; Universitat de Barcelona, Ciència dels Materials i Química Física Tomas, Maria Santos; Universitat Politecnica de Catalunya, Architecture Technology Rubio Martinez, Jaime; Universitat de Barcelona, Ciència dels Materials i Química Física

SCHOLARONE™
Manuscripts

1
2
3 **Fragment Dissolved Molecular Dynamics: A systematic and efficient**
4 **method to locate binding sites**
5
6

7 Cristian Privat¹, José M. Granadino-Roldán², Jordi Bonet¹, Maria Santos Tomas³, Jaime
8 Rubio-Martinez^{*1}
9

10
11 ¹*Departament de Ciència dels Materials i Química Física, Universitat de Barcelona (UB) and*
12 *the Institut de Química Teòrica i Computacional (IQTCUB), Martí i Franqués 1, 08028*
13 *Barcelona, Spain*

14
15 ²*Departamento de Química Física y Analítica, Facultad de Ciencias Experimentales,*
16 *Universidad de Jaén, Campus "Las Lagunillas" s/n, 23071, Jaén, Spain*

17
18 ³*Department of Architecture Technology, Universitat Politècnica de Catalunya, Av. Diagonal*
19 *649, 08028 Barcelona, Spain*
20
21
22
23
24
25

26
27 * Correspondence and requests for materials should be addressed to J.R.M. (email:
28 jaime.rubio@ub.edu)
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Fragment-based drug discovery (FBDD) has been popular in the last decade, but some drawbacks, such as protein denaturation or ligand aggregation, have not yet clearly overcome in the framework of biomolecular simulations. In this work a systematic and semi-automatic method is presented as a novel proposal, named fragment dissolved Molecular Dynamics (fdMD), to improve research in future FBDD projects. Our method employs simulation boxes of solvated small fragments, adding a repulsive Lennard-Jones potential term to avoid aggregation, which can be easily used to solvate the object of interest. This method has the advantage of solvating the target with a low number of ligands, thus preventing this way denaturation of the target, while simultaneously generating a database of ligand-solvated boxes that can be used with other targets. A number of scripts are made available to analyze the results and obtain the descriptors proposed as a means of trustfully discard spurious binding sites. To test our method, four sets of different complexity have been solvated with ligand boxes and four molecular dynamics runs of 200 ns length have been run for each system, which have been extended up to 1 μ s when needed. The reported results point that the selected number of replicas are enough to identify the correct binding sites irrespective of the initial structure, even in the case of proteins having several close binding sites for the same ligand. Among the proposed descriptors, average MMGBSA and average K_{DEEP} energies emerge as the most robust ones.

1. INTRODUCTION

High-throughput screening (HTS) has been a very important approach for drug discovery since 1990s, thanks to which drugs to treat cancer (Sorafenib), HIV (Tripanavir) or diabetes (Sitagliptin), among many others, have been developed¹. However, even using multimillion compound libraries the awareness of the huge chemical space to be explored, estimated at 10^{63} possible small drug-like molecules,² allowed fragment-based drug discovery (FBDD) to emerge as a new strategy to overcome some of the HTS drawbacks. Thus, with a fragment library consisting of 1000 to 5000 compounds adhered to the “rule of three”,³ a reasonable number of hits can be obtained without overwhelming screening methods. FBDD has nowadays become an established technique which has allowed the development of more than 30 drug candidates.⁴

FBDD has also been approached from a theoretical perspective. Many of these approaches are inspired by Dagmar Ringe’s multiple solvent crystal structures (MSCS) method,⁵ which involves solving the crystal structure of the protein in the presence of different organic cosolvents, so that locations of the protein in which these cosolvents repeatedly appear are correlated with binding sites, hotspots and even allosteric sites. These methods have evolved from using static structures to represent the protein, like GRID,⁶ MPS,⁷ or FTMap,⁸ to molecular dynamics (MD) based approaches that allow the conformational adaptation of the protein to the presence of the probes simulating the protein in a solution of water and cosolvent molecules. Among the latter, Barril’s MDmix⁹ was the first to adopt the MD approach, followed by others like SILCS¹⁰ and MixMD.¹¹ One of the important parameters which these methods deal with different approximations is the cosolvent concentration, because on the one hand a high cosolvent concentration can lead to protein denaturation,¹² and on the other it would need the use of artificial repulsion terms in order to prevent aggregation of hydrophobic cosolvents.¹⁰ Different strategies have been presented to overcome these limitations. For example, MixMD avoids these problems with the use of miscible cosolvents at very low concentrations.

Another theoretical approach in the FBDD framework is that followed by Fabritiis and coworkers, in which a fragment is randomly placed in a solvated box and multiple short MD runs are performed to obtain ensembles up to 100 μ s,^{13, 14} with the final goal of predicting the binding site and binding mode of a set of fragments. The use of very long MD simulations has also been proven useful to predict the binding sites and binding modes of ligands. Thus, 10 μ s or longer MD simulations with a box containing 6 dasatinib molecules,¹⁵ or 1 μ s or longer simulations with a box containing 10 GPCR ligands¹⁶ resulted in spontaneous binding events. Another approach aimed at finding the binding site and binding pose of full-sized ligands using a high number of copies of the ligand in the box (64-96), which required the use of a repulsive term in the form of a Lennard-Jones potential, has recently been published.¹⁷

In this study we propose a MD based systematic and semi-automatic procedure aimed at finding the binding site of a small ligand without any previous knowledge of which is the real/experimental binding site of that ligand with its target protein, thus allowing its conformational adaptation. We think that our method could be of use in the initial stages of a FBDD project, in which the affinities and the binding sites of a database of fragments to a target protein are obtained to, later on, grow the most promising fragments to finally obtain a drug, for which having information of the binding site is essential. The method,

1
2
3 which we call fdMD (**f**ragment **d**issolved **M**olecular **D**ynamics), uses a previously
4 equilibrated box containing one copy of the ligand solvated with TIP3P water
5 molecules.¹⁸ Thus, solvation of different proteins with this pre-equilibrated box renders
6 simulation systems containing from 6 to 26 ligand molecules, which prevents the problem
7 of protein denaturation. Besides, the use of a selected set of descriptors allows to rank the
8 located binding sites, select the correct binding site as the best positioned, and discard
9 spurious binding sites using four MD simulation of 200 ns length.¹⁹
10
11

12 2. METHODS

13
14
15 **2.1. fdMD Flowchart.** The fdMD procedure is summarized in Figure 1. In *step 1*, the
16 ligand is meticulously prepared for the following calculations with Amber18.²⁰ Next, the
17 ligand is solvated with TIP3P¹⁸ water molecules using the LEaP module of Amber18,²⁰
18 generating a solvated ligand box. Then, the solvated ligand box is subjected to
19 minimization followed by a MD run, and the final snapshot of that MD run is exported to
20 build the solvation box (a more detailed explanation is described in the SI). Then, each
21 protein, prepared in *step 2*, is solvated with LEaP in *step 3* using the equilibrated ligand
22 box obtained in *step 1* (a more detailed explanation is available in the SI) and after a
23 minimization and equilibration protocol, four independent MD runs of 200 ns length
24 minimum (*step 4*) are computed. Each trajectory is analyzed using four scripts. The first
25 script (fdMD_OneTraj.cmd), which corresponds to *step 5* in Figure 1, strips water
26 molecules and counterions from the trajectory, divides the global trajectory into N
27 protein-ligand trajectories (each one containing the protein and only one ligand), being N
28 the number of ligands in the simulation box, and writes the last snapshot of each of the N
29 trajectories in PDB format. The remaining scripts, devoted to the analysis, correspond to
30 *step 6* in Figure 1. Thus, the second script (fdMD_ReactiveTraj.py) decides which of the
31 N trajectories are reactive, analyzing if the ligand is interacting with a possible binding
32 site during a pre-established simulation time at the end of the trajectory (an extended
33 discussion about our concepts of *reactive trajectory* and *binding site* is described in the
34 SI). Scripts 3 and 4 (fdMD_MMGBSA_Send.cmd and fdMD_MMGBSA_Analyze.py)
35 generate the inputs needed to run a MMGBSA²¹ calculation and to extract and graph data
36 from those calculations. The user then visualizes the possible binding sites occupied and
37 can tabulate, for each trajectory, how much time the ligand remains in each possible
38 binding site (the residence time, RT) and the MMGBSA binding energy for the last 20 ns
39 of trajectory. Using the last snapshot of each trajectory one can also tabulate the binding
40 energy calculated with K_{DEEP} ,²² a fast machine-learning approach for predicting binding
41 affinities using 3D-convolutional neural networks.
42
43
44
45
46

47 We propose to rank the possible binding sites using a consensus of six descriptors. Two
48 of these descriptors (average MMGBSA binding energy and average K_{DEEP} binding
49 energy) consider the number of trajectories in which a ligand interacts with a particular
50 binding pocket (number of reactive trajectories, nr): the average MMGBSA binding
51 energy and average K_{DEEP} binding energy are obtained from the average value of the total
52 number of trajectories (nt). If there is no interaction of the ligands with that binding site
53 in a trajectory, the value of the magnitude to be averaged in the calculation of the
54 descriptor ($\Delta G_i^{\text{MMGBSA}}$ or $\Delta G_i^{\text{KDEEP}}$, see equation 1, which shows how to obtain $\Delta G_{\text{average}}^{\text{MMGBSA}}$)
55 would be zero. Thus, the proposed descriptors are average MMGBSA binding energy,
56 average K_{DEEP} binding energy, best MMGBSA binding energy, best K_{DEEP} binding
57 energy, number of reactive trajectories (i.e. how many of the four trajectories end up with
58 one ligand interacting with the analyzed binding pocket), and longest RT. After analyzing
59
60

1
2
3 all found binding sites, the one with the best descriptors is chosen as a reference. Then,
4 the descriptors for the remaining binding sites are compared to those of the best one. We
5 propose as a consensus criterium to consider a binding site as false if more than three
6 descriptors are worse than those of the chosen as best binding site. However, a binding
7 site can be considered as probable if two or more descriptors are better than those of the
8 best binding site. When more than one probable binding sites are encountered, the MD
9 simulation time should be increased, the descriptors re-calculated, and the analysis
10 repeated (a detailed description of the application of this protocol to solve any possible
11 situation is described in the SI). It is important to note that there may be more than one
12 experimental binding site. The scripts used, together with a toy example, can be
13 downloaded at <https://github.com/DrugDesignUBUJA/fdMD>.
14
15

$$\Delta G_{average}^{MMGBSA} = \frac{\sum_{i=1}^{nr} \Delta G_i^{MMGBSA}}{nt} \text{ with } nt = 4 \quad (1)$$

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

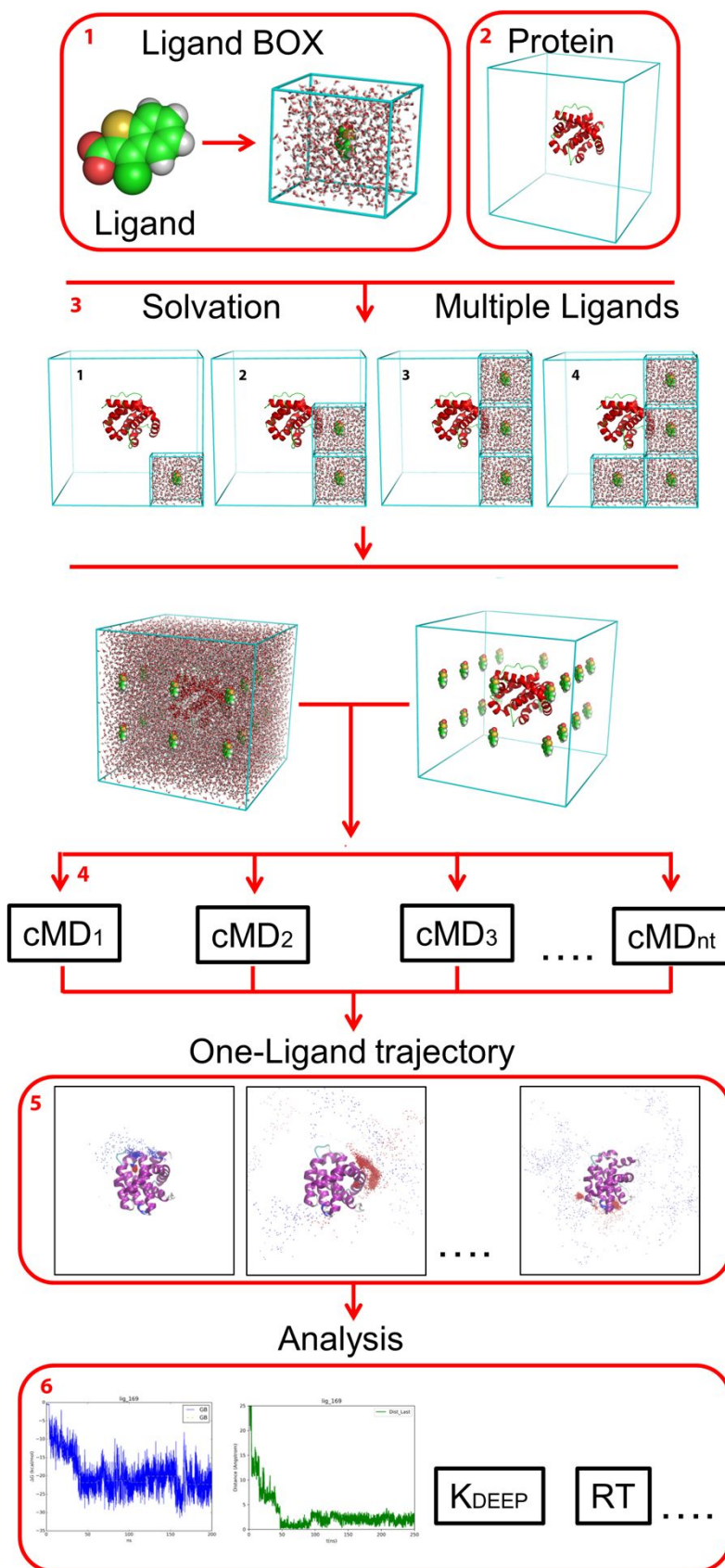
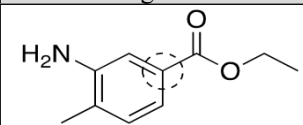
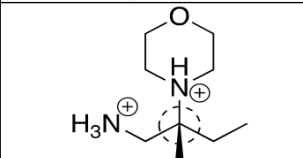
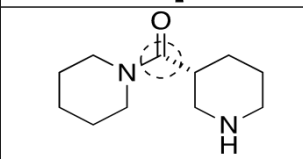
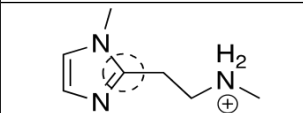
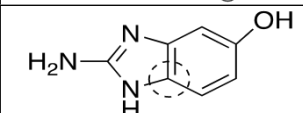
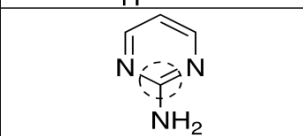
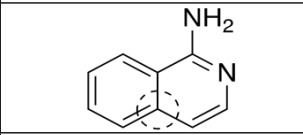
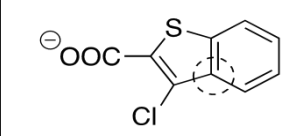
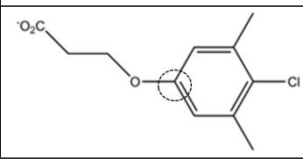
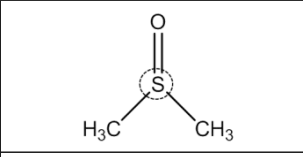
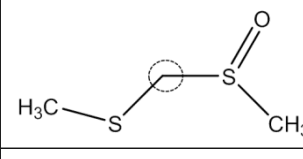
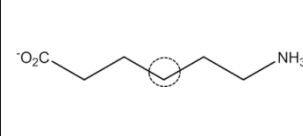


Figure 1. The fdMD flowchart. Each step is labelled in red.

1
2
3 **2.2. Test Systems.** Four different protein-ligand sets were selected for this study (see
4 Table 1), which are intended to show examples of increasing difficulty. Set I is designed
5 to check if the fdMD methodology is robust enough to find the correct binding site
6 irrespective of the initial protein structure (therefore, we compare the protein
7 experimental structure bound to that ligand or another structure of the same protein bound
8 to another ligand). Thus, we selected four ligands of the 38 for which G. Klebe and
9 collaborators²³ obtained an experimental X-ray structure bound to endothiapsin. This
10 protein exhibits a large, solvent exposed, pocket that the authors divided into nine sub
11 pockets and labeled according to the Schechter and Berger nomenclature.²⁴ We have
12 chosen four ligands, namely f031, f035, f207 and f240 (their structures are shown in Table
13 1), so that they span the protein pocket. For each ligand, the fdMD methodology has been
14 essayed using, as stated before, two different initial protein structures. In the first one the
15 experimental structure of the ligand bound to that protein (labelled with the name of the
16 ligand) was used, while in the second one the protein structure was the *holo* form bound
17 to the randomly selected f278 ligand, with PDB code 5DR1²³ (labelled as
18 NameLigand_5DR1).
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Protein-ligand systems and number of ligands of each fdMD simulation.

	Ligand ^a	Protein	Label	PDB ID	Ligands ^b
SET I		Endothiapepsin	f031	5DPZ ²³	9
			f031_5DR1	5DR1 ²³	
			f035	4Y3W ²³	12
			f035_5DR1	5DR1 ²³	
			f207	4Y3T ²³	12
			f207_5DR1	5DR1 ²³	
	f240	4YD6 ²³	8		
	f240_5DR1	5DR1 ²³	12		
SET II		Urokinase	1fv9	1FV9 ²⁵	26
		Hsp90	2jjc	2JJC ²⁶	25
		BACE-1	2ohk	2OHK ²⁷	25
SET III		MCL-1	class1	4OQ5 ²⁸	20
			class2		26
SET IV		FKBP	dms	1D6O ²⁹	8
			dss		6
		HPK4	aca	1PMK ³⁰	8

^a The dashed line circle indicates the C99/S99 atom (see below)^b Number of ligands in the simulation box.

1
2
3 Set II evaluates the efficiency of the fdMD methodology in three different proteins (see
4 Table 1), each one using the corresponding ligand of the X-ray structure and thus showing
5 each protein in its *holo* form. The experimental binding sites of set II are much less solvent
6 exposed than those of endothiapepsin, supposing a greater challenge for our fdMD
7 method. The systems were selected from a set of 12 protein-fragment complexes reported
8 by M. Congreve et al.,²⁶ set also used in a theoretical article that essayed a combined
9 Multiple Copy Simultaneous Search (MCSS) and MMGBSA approach,³¹ avoiding the
10 existence of cofactors in the structures.
11
12

13 Set III deals with the usual situation in a FBDD project, in which the available protein
14 structure is that in its *apo* form. Besides, set III also tests the performance of our
15 methodology to detect the correct binding site for a system composed of bigger ligands.
16 Actually, one of the drawbacks of the fdMD methodology comes from the fact that the
17 repulsion term used to prevent aggregation of ligands (see the Simulation Details section)
18 limits the size of the ligand to be studied, because the cut-off distance for the van der
19 Waals interactions is not big enough. While this limitation is intended to be fixed, in this
20 work we overcome this restriction by splitting the ligand into two parts. Thus, set III is
21 composed of two ligands, 3-chlorobenzo[*b*]thiophene-2-carboxylate (class1) and 3-(4-
22 chloro-3,5-dimethylphenoxy)propanoate (class2), which individually bind myeloid cell
23 leukemia 1 (MCL-1) with K_i of 131 μM and 60 μM , respectively. No experimental
24 structure is available for these two fragments in complex with MCL-1. However, when
25 linked together these two fragments render a compound with $K_i = 0.32 \mu\text{M}$,³² for which
26 an X-Ray structure is available (PDB code 4HW3),³² although the target essayed in this
27 study will use the *apo* form of the MCL-1 protein with PDB code 4OQ5.²⁸
28
29
30
31

32 Finally, set IV is aimed at comparing and discussing the performance of fdMD in three
33 systems studied by ColDock,¹⁷ a recently published method that determines protein-
34 ligand complexes and also uses a repulsive term in the form of a Lennard-Jones potential.
35 All the selected complexes start with the *apo* structure of the protein (see Table 1) even
36 though their corresponding *holo* structures are available. One of these three ligands is
37 dimethylsulphoxide (dms), highly miscible in water and with a $K_i = 20 \text{ mM}$ with the
38 FK506 binding protein (FKBP).³³ The other two fragments, methyl sulphanyl-methyl
39 sulphoxide (dss) and ϵ -aminocaproic acid (aca), present $K_i = 250 \mu\text{M}$ with the FKBP
40 protein³³ and $\text{IC}_{50} = 40\text{-}105 \mu\text{M}$ with the human plasminogen kringle 4³⁴ (HPK4),
41 respectively.
42
43

44 **2.3. Protein Preparation.** Each protein structure was downloaded from the Protein Data
45 Bank³⁵ (PDB codes are shown in Table 1). Water molecules and crystallized ligands were
46 removed, disulfide bonds were created when needed, and residues were protonated with
47 the Protein Preparation Wizard³⁶ of Maestro 2016-2³⁷ at the experimental pH, when
48 available. Missing residues in the 2OHK PDB were included through homology
49 modelling with the X-ray structure of memapsin 2³⁸ (PDB code 1FKN) using the Prime³⁹,
50 ⁴⁰ module of Maestro 2016-2.³⁷ Each protein structure was then loaded into the LEaP
51 module of the Amber18²⁰ suite, where counterions, if needed, were added, and each
52 protein was solvated using TIP3P water molecules¹⁸ with a minimum distance from the
53 edge of the box of 15 Å, and removing those water molecules closer than 2.0 Å from any
54 protein atom. The ff14SB⁴¹ force field was used and the solvated protein was minimized
55 with pmemd¹⁹ for 10000 steepest descent (SD) steps with a cut-off of 10 Å for non-
56 bonded interactions, being electrostatic interactions modelled with the particle-mesh
57 Ewald method⁴². The *apo* structure of MCL-1 used for set III was obtained with a specific
58
59
60

1
2
3 procedure. Starting from the X-ray structure with PDB code 4OQ5, which was prepared
4 in the same way, the minimized, ligand free, 4OQ5 structure, was subjected to a heating
5 and equilibration protocol with the CUDA version of pmemd⁴³⁻⁴⁵. All bonds involving
6 hydrogen atoms were constrained to their equilibrium value using the SHAKE
7 algorithm.⁴⁶ A two steps preparation protocol with an integration time step of 2 fs was
8 used which included increasing the temperature of the system to 300 K at a constant rate
9 of 1.5 K ps⁻¹ during 200 ps in the NVT ensemble and 500 ps at 300 K in the NPT ensemble
10 to adjust density. After 200 ns of MD 300 K in the NVT ensemble, a clustering analysis
11 with cpptraj²⁰ allowed to extract the representative of the most populated cluster as an
12 *apo* structure for MCL-1.
13
14

15
16 **2.4. fdMD Simulations.** Each ligand was prepared with Maestro 2016-2³⁷ and optimized
17 with semiempirical Austin Model 1 (AM1).⁴⁷ Parameters and charges for the ligands were
18 then obtained using Antechamber²⁰ with generalized Amber force field (GAFF2)
19 parameters,⁴⁸ whilst partial charges were derived through the AM1-BCC method.^{49, 50}
20 One of the central carbon (sulphur in the dmsol system) atoms names (indicated in each
21 ligand structure shown in Table 1 with a dashed circle) was changed to C99/S99 to later
22 apply a repulsion term needed to avoid ligand aggregation. These parameters were loaded
23 into LEaP, where counterions, if needed, were added, and each ligand was solvated using
24 TIP3P¹⁸ water molecules, as done during the protein preparation step. Each box was then
25 subjected to 10000 SD minimization steps and the same heating and equilibration
26 protocol used for MCL-1. The production run consisted of 10 ns of MD, and the last
27 snapshot of the production run was used as a solvation box for the preparation of each of
28 the systems studied (a more detailed explanation of the procedure used, together with
29 information about the final boxes, is described in the SI).
30
31
32

33 The minimized structure of each of the proteins studied was used, without water and
34 counterions, as a starting point for fdMD simulations (see Protein preparation). Thus,
35 each protein was solvated with the corresponding equilibrated ligand solvation box. In
36 this process, any molecule in the ligand solvation box closer than 1.0 Å from any protein
37 atom was removed, and the minimum distance from the edge of the box was chosen,
38 depending on the size of the ligand and the shape of the protein, so that a reasonable
39 number of ligands were placed in the simulation box, thus using values from 11 to 20 Å
40 (a more detailed explanation is described in the SI). Finally, counterions were added when
41 needed. The ff14SB⁴¹ force field was used for subsequent calculations. Besides,
42 ParmEd²⁰ was used to modify the topology in order to use hydrogen mass repartitioning⁵¹
43 (HMR) and to add a Lennard-Jones repulsion potential, setting to zero the attractive part
44 ($B_{C99/S99-C99/S99} = 0$), between C99/S99 atoms of the ligands, with $\epsilon = -0.01$ kcal·mol⁻¹ and
45 $R_{\min} = 23.0$ Å.
46
47
48

49 Each system was minimized and equilibrated using the previously described protocol and
50 parameters, with the exception of the cut-off for non-bonded interactions which was
51 increased to 14 Å. Four independent production runs of 200 ns length with an integration
52 time of 4 fs were then used to select the reactive trajectories and score them according to
53 the proposed descriptors.
54
55

56 **2.5. fdMD Analysis: Selection and Scoring of Reactive Trajectories.** Each of the
57 stripped trajectories, which contain the protein and one of the ligands from a ligand
58 solvation box used to solvate the protein, were obtained with the fdMD_OneTraj.cmd
59 script. The fdMD_ReactiveTraj.py script analyzed these trajectories by means of the
60

1
2
3 Euclidean distance for atom C99/S99, measured between the last snapshot and all
4 previous ones ($\text{dis}(C99_{\text{last}}-C99_t)/\text{dis}(S99_{\text{last}}-S99_t)$, being $C99_{\text{last}}/S99_{\text{last}}$ the position of the
5 C99/S99 atom in the last snapshot of the trajectory and $C99_t/S99_t$ the position of the
6 C99/S99 atom at each time t of the trajectory. Thus, a trajectory is considered reactive if
7 (i) the C99/S99 atom has at least one protein atom at a distance below a predefined cut-
8 off (5 Å in this work) in the last snapshot and (ii) this distance remains below the cut-off
9 at least during the 90% of snapshots during the last 20 ns of the simulation time. A picture
10 of the evolution of the ($\text{dis}(C99_{\text{last}}-C99_t)/\text{dis}(S99_{\text{last}}-S99_t)$) along the simulation time is
11 plotted (see the second picture of *step 6* in Figure 1) by the script.
12
13

14 Furthermore, another script is outputted to delete the files related to non-reactive
15 trajectories, together with a PDB file of each last snapshot to be used to obtain K_{DEEP}^{22}
16 binding energies. The MMGBSA calculations which are automatically sent for each of
17 the reactive trajectories (script `fdMD_MMGBSA_Send.cmd`) used no entropic
18 contributions, with the modified GB model by Onufriev et al.⁵² (`igb=2`) and all remaining
19 parameters set at their default values. Finally, `fdMD_MMGBSA_Analize.py` was
20 instructed to calculate the average binding energy in intervals of 20 ns for all the reactive
21 trajectories. Also, a picture of the evolution of the $\Delta G_{\text{binding}}$ obtained by the MMGBSA
22 method along the simulation time is generated by this script (see the first picture of *step*
23 *6* in Figure 1).
24
25
26

27 3. RESULTS AND DISCUSSION

28
29 **3.1. Set I: Proving the Robustness of fdMD with the Endothiapepsin Protein.** Results
30 for set I are summarized in Tables 2 and S3. In general terms, no more than 2 pockets are
31 found in each of the systems studied despite using 8 to 12 ligands in 4 different MD runs.
32 Besides, results are equally good irrespective of the initial protein structure, indicating
33 that the method can avoid any possible bias due to the conformation of the receptor at the
34 beginning of the simulation.
35
36

37 For the f031 (and f031_5DR1) system, two possible binding sites (pockets S1' and S6)
38 are found, being S1' the best one according to the descriptors. Both of them, interestingly,
39 exhibit a hydrophobic character. Among these two binding sites, Pocket S6 would be
40 easily discarded as false positive using the proposed descriptors, which always point that
41 S1' is better with the exception of the best K_{DEEP} binding energy (that is, just one
42 descriptor is better for S6 than for S1'). The f035 and f035_5DR1 systems have 2
43 experimental binding sites, pockets S3 and S6, whose proximity prevent fdMD from
44 obtaining reactive trajectories in both pockets for the same MD run (see Table S3).
45 Precisely, the low time the ligand is interacting with pocket S6 in MD run 4 of the f035
46 system (just 20 ns), is due to the fact that the ligand is oscillating between both pockets
47 (see video as Web Enhanced Object). F207 experimentally binds to three pockets: the
48 catalytic dyad, pocket S6 and pocket S2. Two of these three pockets (the catalytic dyad
49 and pocket S6) are predicted by fdMD. This fact can easily be explained looking the
50 spatial distribution of these three pockets; pocket S2 is located between the catalytic dyad
51 and pocket S6. Thus, the two extremes are far enough to appear interacting with their
52 ligand even in the same trajectory (Table S3), but pocket S2 is too close to the other
53 two pockets to appear in the same trajectory interacting with the ligand when pocket S6
54 or the catalytic dyad have already a ligand nearby. In addition, the small binding affinity
55 of this ligand for pocket S2 probably prevents it to arise alone. No false positives are
56 predicted and the descriptors, with the exception of best K_{DEEP} binding energy for
57
58
59
60

1
2
3 f207_5DR1, suggest that the catalytic dyad is the preferred binding site. F240 exhibits a
4 false positive (pocket S6) in one of the MD runs, while no false positives appear for the
5 f240_5DR1 system. This false positive can be easily discarded after analyzing the
6 proposed descriptors, as all its descriptors are worse than those of the best binding pocket
7 (pocket S1). Figures 2 and S1 illustrate experimental and predicted poses using the natural
8 or 5DR1 receptor, respectively. These figures allow concluding that, although the binding
9 sites are correctly located, the experimental binding mode is not always found.
10
11

12 It should be noted that for these particular systems, averaged MMGBSA, best MMGBSA
13 and average K_{DEEP} values appear as good global descriptors for this set. They clearly
14 indicate that the best binding site and the energetic order for the different ligands is
15 maintained irrespective of the receptor. The only case where the value of these descriptors
16 is very different for the two receptors is pocket S3 of the f035 system, because different
17 binding poses are found for each receptor structure, corresponding the best MMGBSA
18 value to the binding site closer to the experimental structure (compare Figure 2b and
19 Figure S1b).
20
21

22 Finally, in order to test the stability of our method with respect to the simulation length,
23 we extended the four MD runs of two of the converged systems, f031_5DR1 and
24 f035_5DR1, up to 300 ns length. In both calculations the worst binding site disappears
25 and the one closer to the experimental structure remains (see Table S4, S5 and Figure
26 S2).
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2. Descriptors calculated for reactive trajectories of set I. Pockets are defined in reference 23. MMGBSA and K_{DEEP} energies in kcal mol⁻¹.

	System ^a	Binding site	Ave. MMGBSA ^b	Ave. K_{DEEP} ^c	Best MMGBSA	Best K_{DEEP}	N. react. ^d	Best RT ^e
Natural Receptor	f031	Pocket S1'	-12.2	-6.0	-15.1	-6.4	4	180
		Pocket S6	-5.2	-3.5	-10.6	-7.2	2	75
	f035	Pocket S3	-6.1	-1.4	-24.4	-5.5	1	120
		Pocket S6	-2.9	-1.1	-11.5	-4.2	1	20
	f207	Dyad ^f	-10.3	-4.3	-18.9	-6.1	3	180
		Pocket S6	-5.5	-2.5	-12.2	-5.2	2	80
f240	Pocket S1	-16.9	-2.9	-18.3	-5.9	2	140	
SDR1 Receptor	f031_5DR1	Pocket S1'	-11.9	-5.2	-12.4	-6.2	4	180
		Pocket S6	-2.7	-1.6	-10.7	-6.4	1	20
	f035_5DR1	Pocket S3	-22.0	-4.3	-31.6	-6.3	3	170
		Pocket S6	-3.9	-1.3	-15.7	-5.2	1	180
	f207_5DR1	Dyad ^f	-15.3	-5.3	-18.5	-5.9	4	170
		Pocket S6	-2.9	-1.4	-11.4	-6.2	1	40
f240_5DR1	Pocket S1	-15.9	-6.2	-17.3	-6.4	4	150	

^a See definition in Table 1. ^b Average MMGBSA binding energy (see Equation 1). ^c Average K_{DEEP} binding energy. ^d Number of reactive trajectories interacting with that binding pocket. ^e In ns. ^f Catalytic dyad. In bold the best value for each descriptor.

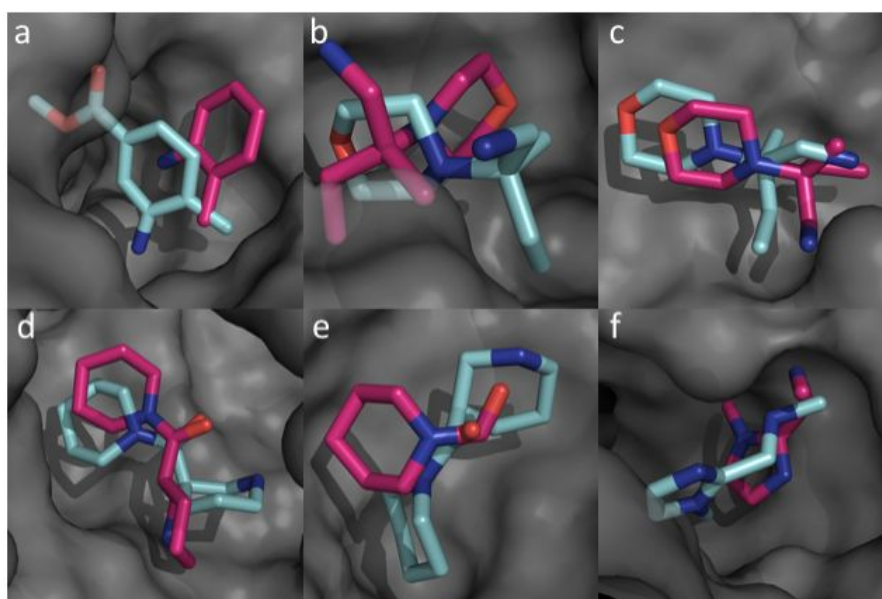


Figure 2. Comparison of experimental (carbon atoms magenta) and predicted (carbon atoms cyan) binding poses for the different systems studied in set I with their experimental receptor. a) f031, pocket S1', b) f035, pocket S3, c) f035, pocket S6, d) f207, catalytic dyad, e) f207, pocket S6, f) f240, pocket S1. Hydrogen atoms are omitted for clarity.

3.2. Set II: Search for Low Solvent-Exposed Pockets. As can be seen in Tables 3 and S7, the results of the fdMD method are system dependent. In the case of the 2jic system, it is easy to decide which binding site is the false positive among the two found. Hence, BS 1, for which the only descriptor favoring this false positive is the best MMGBSA

1
2
3 binding energy, can be discarded. On the other hand, the 1fv9 system presents nine
4 possible binding sites, but in this case BS 1 is the binding site with best descriptors. The
5 experimental pocket should be considered as doubtful, taking into account that two
6 descriptors (average K_{DEEP} binding energy and number of reactive trajectories) are better
7 for it than for BS 1. We increased the simulation time to 300 ns to shed some light on this
8 system. All previous reactive ligands, except for those in BS 1, BS 3 and the experimental
9 one, left their respective binding sites and, at the same time, allowed to easily identify BS
10 1 and BS 3 as spurious binding sites after comparing the descriptors. Now the
11 experimental binding site is the best one, and best K_{DEEP} binding energy, with a small
12 difference of $0.1 \text{ kcal mol}^{-1}$ is the only descriptor better for BS 1 and BS 3 than for the
13 experimental one.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

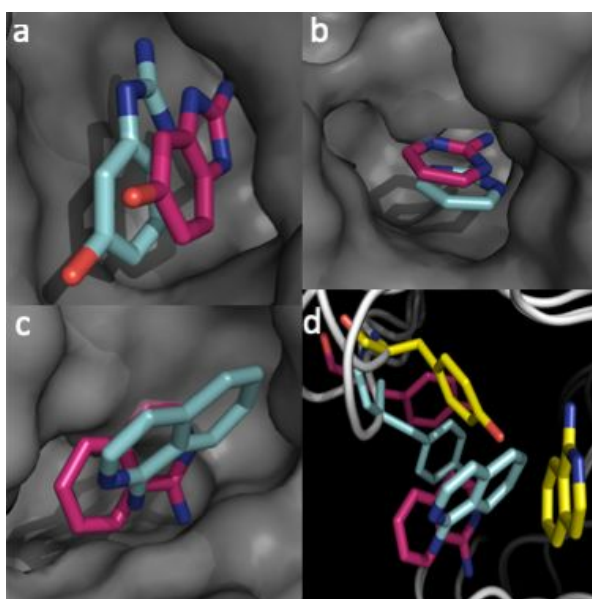
Table 3. Descriptors calculated for reactive trajectories of set II. Pockets are defined in Table S6 and reference 53. MMGBSA and K_{DEEP} energies in kcal mol⁻¹.

Simulation time	System ^a	Binding site	Ave. MMGBSA ^b	Ave. K_{DEEP} ^c	Best MMGBSA	Best K_{DEEP}	N. react. ^d	Best RT ^e
200 ns	1fv9	Experimental	-11.6	-3.7	-17.8	-5.2	3	190
		BS 1	-12.0	-2.7	-25.4	-5.7	2	190
		BS 2	-3.1	-1.4	-12.3	-5.6	1	40
		BS 3	-8.5	-2.4	-17.0	-4.9	2	138
		BS 4	-3.0	-1.3	-12.1	-5.0	1	30
		BS 5	-3.3	-1.3	-13.1	-5.0	1	152
		BS 6	-4.0	-0.8	-16.0	-3.2	1	49
		BS 7	-2.7	-1.4	-10.9	-5.7	1	54
300 ns		Experimental	-15.6	-3.9	-25.4	-5.5	3	290
		BS 1	-5.7	-1.4	-22.7	-5.6	1	290
		BS 3	-8.4	-2.8	-16.6	-5.6	2	238
200 ns	2jjc	Experimental	-6.9	-3.6	-11.0	-4.9	3	190
		BS 1	-4.6	-1.0	-18.2	-4.0	1	170
200 ns	2ohk	S1	-5.85	-2.6	-11.8	-5.2	2	161
		BS 1	-2.48	-1.1	-9.9	-4.4	1	39
		BS 2	-2.00	-0.9	-8.0	-3.6	1	20
		BS 3	-3.60	-0.9	-14.4	-3.6	1	180
		BS 4	-9.28	-3.6	-13.8	-5.6	3	141
		BS 5	-1.68	-1.4	-6.7	-5.6	1	62
300 ns		BS 6	-4.35	-2.3	-8.8	-5.5	2	47
		S1	-5.3	-2.8	-11.0	-5.7	2	261
		BS 3	-4.1	-1.1	-16.4	-4.5	1	280
1 μ s		BS 4	-9.6	-3.7	-12.8	-5.6	3	241
		Experimental	-2.5	-1.5	-9.9	-5.9	1	28
		BS 4	-3.3	-1.1	-13.2	-4.2	1	829

^a See definition in Table 1. ^b Average MMGBSA binding energy (see Equation 1). ^c Average K_{DEEP} binding energy. ^d Number of reactive trajectories interacting with that binding pocket. ^e In ns. In bold the best value for each descriptor.

Probably, the 2ohk system (BACE-1 protein) is the most complicated to interpret because of the binding mode of the ligand. Seven possible binding sites are found, none of them corresponding to the experimental one, although two reactive trajectories show the ligand interacting with an already reported hydrophobic pocket named S1.⁵³ It is very interesting to realize that, newly synthesized, bigger BACE-1 inhibitors actually interact both with S1 and the catalytic aspartate dyad,⁵⁴⁻⁶⁰ with which 3-chlorobenzo[*b*]thiophene-2-carboxylate interacts. A simulation time of 200 ns would allow to conclude that the best binding site is BS 4, and only BS 3, for which the best MMPBSA binding energy and the residence time are better, could be considered doubtful. Then, an increase in 100 ns of each of the MD runs implies that the ligands in all binding sites apart from S1, BS 3 and BS 4 move away from their corresponding binding sites, and the experimental binding site would not still been found. Another increase in the simulation time from 300 ns to 1 μ s allows 3-chlorobenzo[*b*]thiophene-2-carboxylate in MD run 4 to move out of the S1 pocket (see Table S3), while that in MD run 2 finds the experimental binding site.

1
2
3
4 A change in the conformation of the side chain of TYR71 plays a key role in the binding
5 mechanism of the ligand (see Figure 3d). The conformation adopted by this TYR71,
6 which is not seen in any *holo* nor *apo* BACE-1 experimental structure, blocks the path to
7 find the experimental binding site and can explain the need for a higher simulation time
8 or enhanced sampling techniques when dealing with cryptic pockets. It is worth to
9 mention that for the 1fv9 and 2jjc systems the correct orientation for the ligands in their
10 corresponding pocket is found (see Figures 3a and 3b), and negligible differences in the
11 binding pose are seen in 1fv9 when going from 200 ns to 300 ns in the simulation time.
12 Contrarily, the right orientation is not achieved for the 2ohk system even with 1 μ s of
13 MD (see Figure 3d).
14
15



16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Figure 3. a-c) Comparison of experimental (carbon atoms magenta) and predicted (carbon atoms cyan) binding poses for the different systems studied in set II. a) 1fv9 after 200 ns, b) 2jjc after 200 ns, c) 2ohk after 1 μ s. d) Comparison of the conformation adopted by TYR71 and position of 1-amino-isoquinoline between the experimental structure (carbon atoms magenta), MD run of 200 ns (carbon atoms yellow) and MD run of 1 μ s (carbon atoms cyan). Hydrogen atoms are omitted for clarity.

3.3. Set III: Apo Form of MCL-1 and Large Size Ligand. This could be considered the most challenging set since the target is used in its *apo* form. As can be seen in Tables 4 and S8, eight possible binding sites are found after 4 MD runs of 200 ns for the class1 system, with the experimental pocket being the best one. In the case of class2, twelve possible binding sites are found, being also the experimental pocket the one with best descriptors (see below). Among these binding pockets, BS 13⁶¹ and BS 14⁶² have already been reported for other MCL-1 inhibitors. Focusing on the class1 system, it can be easily concluded that all the binding sites apart from the experimental one can be discarded according to our criteria. However, although the binding pocket is found, the predicted binding pose needs improvement (see Figure 4), so further 100 ns were calculated. As happened before, an increase in the simulation time implied that some ligands moved away from their binding site, while also the binding pose improved in the experimental binding site. It is important to note that in this set we are using fragments of the real compound, while the complete experimental structure is taken as a reference. Therefore,

1
2
3 some differences are expected in their binding mode as a consequence of differences in
4 size and binding interactions with the protein.
5

6
7 In the case of the class2 system, fdMD finds the experimental binding pocket with only
8 200 ns of molecular dynamics, although the binding pose adopted by this part of the
9 original ligand is similar to the pose found by the fragment of the class1 system (see
10 Figure 4). An increase of 100 ns induces six ligands to move out of their corresponding
11 binding site, and the remaining ones can also be marked as spurious using the proposed
12 descriptors and consensus criterium, since all descriptors are worse for them (BS 4, BS
13 6, BS 10, BS 11 and BS 14). In order to check if more simulation time would allow the
14 class2 system to find the correct binding pose, a further increase in the simulation time to
15 1 μ s was performed. This increase in the simulation time implied that ligands in all
16 binding sites but the experimental one leave their binding sites while 3 reactive
17 trajectories interacting with the experimental class1 binding site remain. As described
18 previously for the 2ohk system, an increase in the simulation time, even to 1 μ s, is not
19 enough to find the correct binding pose. However, for this particular case we can point
20 towards two different hypotheses. The first one suggests the convenience of using
21 enhancing sampling techniques when the final goal is not only to find the binding site but
22 also the right binding pose. A second hypothesis suggests concluding that this fragment
23 exhibits a binding mode different to that of the full ligand. We are currently studying the
24 first hypothesis in order to check whether the fdMD method is capable of providing more
25 accurate results applying enhanced sampling techniques.
26
27
28

29
30 **3.4. Set IV: Comparison with ColDock.** The dmsol system was selected to clearly
31 remark the different approach used by our method with respect to ColDock. As can be
32 seen in Tables 5 and S9, none of the four MD trajectories of 200 ns length was reactive
33 with our proposed criteria. This is because the binding energy of dmsol to the FKBP
34 protein is very small and has a clear dynamic binding. The ligand interacts with a
35 particular binding site of the receptor, remains in this site for a short period of time and
36 goes out to another protein region, and so on continuously. Thus, a statistical approach
37 which counts the number of times that a dmsol molecule interacts with a specific point
38 can be used to detect druggable pockets. This is the philosophy of the cosolvent approach
39 and also the one used by ColDock. Our method does not use a statistical approach, but it
40 requests that the ligand has a reasonable binding energy and thus keeps bonded in a
41 specific binding site for a long period of time. It is important to note that all our four MD
42 runs detect the binding of some dmsol molecules to the experimental binding site, but
43 always leave it (see Figure S3). The results for the other two systems are described in
44 Table 6 (see also Table S9). As it can be seen, for both ligands the method detects only
45 the experimental binding site with good agreement with the corresponding binding pose
46 (Figure 5).
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 4. Descriptors calculated for reactive trajectories of set III. Pockets are defined in Table S2. MMGBSA and K_{DEEP} energies in kcal mol⁻¹.

Simulation time	System ^a	Binding site	Ave. MMGBSA ^b	Ave. K_{DEEP} ^c	Best MMGBSA	Best K_{DEEP}	N. react. ^d	Best RT ^e
200 ns	class1	Experimental	-21.1	-5.8	-22.3	-6.2	4	195
		BS 1	-2.4	-1.1	-9.7	-4.5	1	25
		BS 2	-7.8	-2.5	-18.7	-5.2	2	60
		BS 3	-6.9	-2.1	-14.1	-4.2	2	50
		BS 4	-16.4	-3.6	-17.4	-3.8	4	140
		BS 5	-7.7	-1.9	-17.6	-4.1	2	60
		BS 6	-6.0	-1.2	-24.0	-4.9	1	100
300 ns		Experimental	-23.1	-5.6	-24.3	-6.4	4	295
		BS 2	-3.1	-1.3	-12.4	-5.0	1	160
		BS 4	-17.7	-3.6	-21.3	-4.0	4	260
		BS 6	-7.0	-1.0	-28.1	-3.8	1	200
200 ns	class2	Experimental	-19.5	-4.6	-28.8	-6.3	3	175
		BS 4	-11.9	-1.8	-24.9	-3.8	2	33
		BS 5	-8.2	-1.6	-19.9	-3.2	2	135
		BS 6	-11.6	-2.1	-31.7	-4.8	2	192
		BS 8	-3.7	-1.1	-14.6	-4.3	1	60
		BS 9	-3.7	-1.0	-14.6	-4.1	1	25
		BS 10	-9.0	-2.3	-20.2	-4.7	2	110
		BS 11	-5.2	-0.7	-20.6	-2.9	1	70
		BS 12	-7.1	-1.3	-28.3	-5.0	1	110
300 ns		BS 13	-4.2	-1.2	-16.6	-4.6	1	20
		BS 14	-3.8	-0.9	-15.3	-3.7	1	122
		BS 15	-3.1	-1.0	-12.4	-3.8	1	30
		Experimental	-18.5	-3.9	-29.8	-6.0	3	275
		BS 4	-6.0	-1.0	-24.0	-3.9	1	133
1 μ s		BS 6	-7.3	-1.0	-29.0	-3.9	1	292
		BS 10	-4.2	-1.0	-16.6	-4.1	1	178
		BS 11	-5.1	-0.8	-20.2	-3.2	1	170
		BS 14	-4.2	-0.9	-16.9	-3.6	1	222
1 μ s		Experimental	-15.9	-4.6	-23.2	-6.5	3	975

^a See definition in Table 1. ^b Average MMGBSA binding energy (see Equation 1). ^c Average K_{DEEP} binding energy. ^d Number of reactive trajectories interacting with that binding pocket. ^e In ns. In bold the best value for each descriptor.

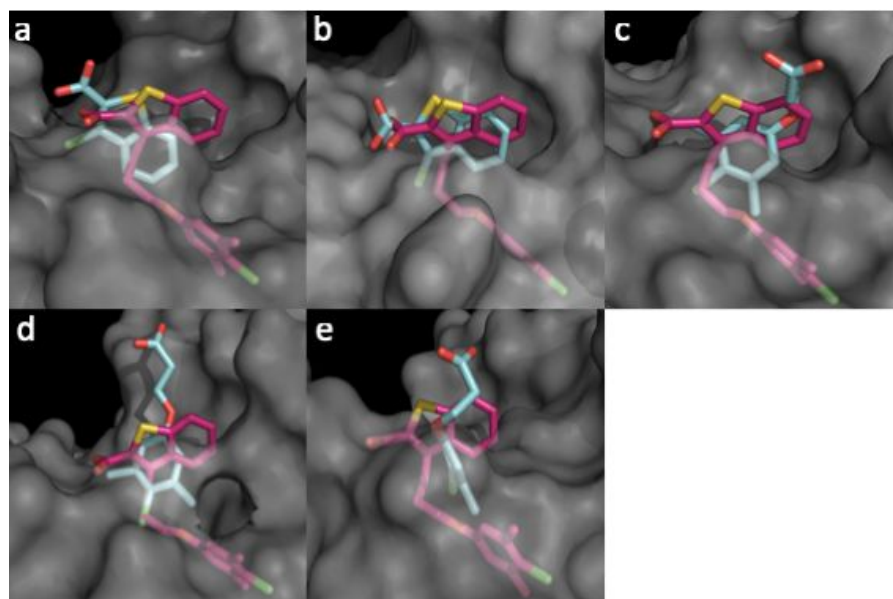


Figure 4. Comparison of experimental (carbon atoms magenta) and predicted (carbon atoms cyan) binding poses for the different systems and MD simulation times studied in set III. a) class1 after 200 ns, b) class1 after 300 ns, c) class2 after 200 ns, d) class2 after 300 ns, e) class2 after 1 μ s. Hydrogen atoms are omitted for clarity.

Table 5. Descriptors calculated for reactive trajectories of set IV. MMGBSA and K_{DEEP} energies in kcal mol⁻¹.

System ^a	Binding site	Ave. MMGBSA _b	Ave. K_{DEEP}^c	Best MMGBSA	Best K_{DEEP}	N. react. ^d	Best RT ^e
dms0							
dss	Experimental	-7.2	-6.0	-14.4	-6.0	2	155
aca	Experimental	-9.2	-5.3	-13.2	-5.9	4	23

^a See definition in Table 1. ^b Average MMGBSA binding energy (see Equation 1). ^c Average K_{DEEP} binding energy. ^d Number of reactive trajectories interacting with that binding pocket. ^e In ns. In bold the best value for each descriptor.

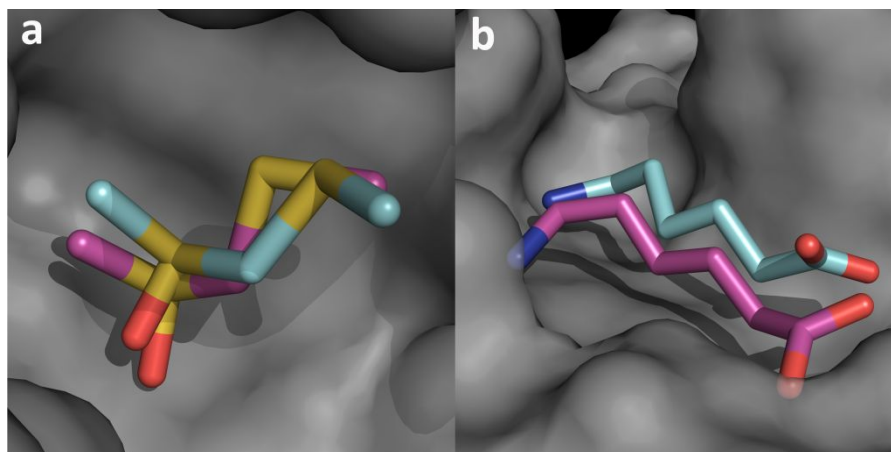


Figure 5. Comparison of experimental (carbon atoms magenta) and predicted (carbon atoms cyan) binding poses for the different reactive systems studied in set IV with their apo receptor. a) dss, b) aca. Hydrogen atoms are omitted for clarity.

4. CONCLUSIONS

This work presents a semi-automated method applicable to any system in the framework of fragment-based drug discovery. Our method, named fragment dissolved Molecular Dynamics (fdMD), is based on the generation of simulation boxes of solvated small fragments that can be easily used to solvate around a target. It is worth to stress that the procedure of obtaining a solvated ligand box is done once and that box can be used later, in an easy way, to solvate any system and to prepare a database of ligand boxes to be used in a fragment-based drug discovery project. fdMD requires a low number of ligands, avoiding this way the risk of denaturation, and modifies the van der Waals interaction of a selected atom within the ligand using a pure repulsion potential to prevent ligand aggregation. Besides, we propose the use of six descriptors as a robust way to discard false binding sites. This work essays the performance of this methodology using four MD runs of 200 ns, which are extended up to 1 μ s when needed.

After studying three sets, which contain a total of five protein systems with different types of binding sites, we conclude that our proposed method finds at least one experimental binding site, which usually corresponds to the most energetically favorable one according to MMGBSA and K_{DEEP} results. Even in the case of ligands with several close binding sites, fdMD is able to recognize them despite the use of a repulsion potential for the ligands as several MD trajectories are used. Also, our results suggest that the effectiveness of the method is not dependent of the initial structure selected for the receptor, which is the usual case in a FBDD project. On the other hand, the number of replicas becomes crucial to discern between correct and false ligand-protein binding sites because it has a direct effect on the statistical average of the descriptors, being 4 MD runs reasonable. The average binding free energies estimated with MMGBSA or K_{DEEP} approaches stand out as reliable descriptors to discard false binding sites, although we must stress that the computational cost of K_{DEEP} is much smaller. Finally, our results suggest that MD runs of 200 ns length are a trustworthy starting point to deal with both exposed and unexposed binding sites reemerging as a new alternative for the identification and optimization of potential fragments for the discovery of novel drugs.

Some aspects mentioned in the discussion are on the table to improve the fdMD method. Future research will focus on the sampling of cryptic pockets or the identification of the correct binding modes by using enhanced-sampling techniques.

5. ACKNOWLEDGEMENTS

This study was supported by The Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR)-Generalitat de Catalunya (2017SGR1033). José M. Granadino-Roldán thanks the University of Jaén (Acción 1 PIUJA 2017-2018) for financial support. We would like to thank the reviewers for helping us to improve the original manuscript.

Supporting Information Available: Detailed description of steps 1, 3 and 6 of the fdMD flowchart. Description of the concept of reactive trajectory and binding site used in fdMD. Tables S1 to S9 and Figures S1 to S3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

6. REFERENCES

1. Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S., Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery* **2011**, *10*, 188.
2. Bohacek, R. S.; McMartin, C.; Guida, W. C., The art and practice of structure-based drug design: A molecular modeling perspective. *Medicinal Research Reviews* **1996**, *16*, 3-50.
3. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H., A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* **2003**, *8*, 876-877.
4. Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H., Twenty years on: the impact of fragments on drug discovery. *Nature Reviews Drug Discovery* **2016**, *15*, 605.
5. Allen, K. N.; Bellamacina, C. R.; Ding, X.; Jeffery, C. J.; Mattos, C.; Petsko, G. A.; Ringe, D., An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. *The Journal of Physical Chemistry* **1996**, *100*, 2605-2611.
6. Goodford, P. J., A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry* **1985**, *28*, 849-857.
7. Carlson, H. A.; Masukawa, K. M.; McCammon, J. A., Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design. *The Journal of Physical Chemistry A* **1999**, *103*, 10213-10219.
8. Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S., Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621-627.
9. Seco, J.; Luque, F. J.; Barril, X., Binding Site Detection and Druggability Index from First Principles. *Journal of Medicinal Chemistry* **2009**, *52*, 2363-2371.
10. Guvench, O.; MacKerell, A. D., Jr., Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Computational Biology* **2009**, *5*, e1000435.
11. Lexa, K. W.; Carlson, H. A., Full Protein Flexibility Is Essential for Proper Hot-Spot Mapping. *Journal of the American Chemical Society* **2011**, *133*, 200-202.
12. Foster, T. J.; MacKerell, A. D., Jr.; Guvench, O., Balancing target flexibility and target denaturation in computational fragment-based inhibitor discovery. *Journal of Computational Chemistry* **2012**, *33*, 1880-91.
13. Ferruz, N.; Harvey, M. J.; Mestres, J.; De Fabritiis, G., Insights from Fragment Hit Binding Assays by Molecular Simulations. *Journal of Chemical Information and Modeling* **2015**, *55*, 2200-2205.
14. Martinez-Rosell, G.; Harvey, M. J.; De Fabritiis, G., Molecular-Simulation-Driven Fragment Screening for the Discovery of New CXCL12 Inhibitors. *Journal of Chemical Information and Modeling* **2018**, *58*, 683-691.
15. Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E., How Does a Drug Molecule Find Its Target Binding Site? *Journal of the American Chemical Society* **2011**, *133*, 9181-9183.
16. Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E., Pathway and mechanism of drug binding to G-protein-coupled

- receptors. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, 108, 13118-13123.
17. Takemura, K.; Sato, C.; Kitao, A., ColDock: Concentrated Ligand Docking with All-Atom Molecular Dynamics Simulation. *The Journal of Physical Chemistry B* **2018**, 122, 7191-7200.
18. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, 79, 926-935.
19. Perez, J. J.; Tomas, M. S.; Rubio-Martinez, J., Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory. *J Chem Inf Model* **2016**, 56, 1950-1962.
20. Case, D.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T.; Duke, R. E.; Ghoreishi, D.; Gohlke, H.; Goetz, A. W.; Green, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A., *Amber 2018*.
21. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012**, 8, 3314-3321.
22. Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G., KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *Journal of Chemical Information and Modeling* **2018**, 58, 287-296.
23. Radeva, N.; Krimmer, S. G.; Stieler, M.; Fu, K.; Wang, X.; Ehrmann, F. R.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Schiebel, J.; Heine, A.; Klebe, G., Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *Journal of Medicinal Chemistry* **2016**, 59, 7561-75.
24. Schechter, I.; Berger, A., On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **1967**, 27, 157-62.
25. Hajduk, P. J.; Boyd, S.; Nettesheim, D.; Nienaber, V.; Severin, J.; Smith, R.; Davidson, D.; Rockway, T.; Fesik, S. W., Identification of Novel Inhibitors of Urokinase via NMR-Based Screening. *Journal of Medicinal Chemistry* **2000**, 43, 3862-3866.
26. Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J., Recent developments in fragment-based drug discovery. *Journal of Medicinal Chemistry* **2008**, 51, 3661-80.
27. Murray, C. W.; Callaghan, O.; Chessari, G.; Cleasby, A.; Congreve, M.; Frederickson, M.; Hartshorn, M. J.; McMenamin, R.; Patel, S.; Wallis, N., Application of Fragment Screening by X-ray Crystallography to β -Secretase. *Journal of Medicinal Chemistry* **2007**, 50, 1116-1123.
28. Petros, A. M.; Swann, S. L.; Song, D.; Swinger, K.; Park, C.; Zhang, H.; Wendt, M. D.; Kunzer, A. R.; Souers, A. J.; Sun, C., Fragment-based discovery of potent inhibitors of the anti-apoptotic MCL-1 protein. *Bioorganic & Medicinal Chemistry Letters* **2014**, 24, 1484-1488.

- 1
2
3 29. Burkhard, P.; Taylor, P.; Walkinshaw, M. D. X-ray structures of small ligand-fkbp
4 complexes provide an estimate for hydrophobic interaction energies. *J. Mol. Biol.*
5 **2000**, 295 (4), 953–962.
6
7 30. Padmanabhan, K.; Wu, T. P.; Ravichandran, K. G.; Tulinsky, A. Kringle-kringle
8 interactions in multimer kringle structures. *Protein Sci.* **1994**, 3, 898-910.
9
10 31. Haider, M. K.; Bertrand, H.-O.; Hubbard, R. E., Predicting fragment binding
11 poses using a combined MCSS MM-GBSA approach. *Journal of Chemical Information*
12 *and Modeling* **2011**, 51, 1092-105.
13
14 32. Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Burke, J. P.; Garcia-Barrantes, P.
15 M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W., Discovery of
16 potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and
17 structure-based design. *Journal of Medicinal Chemistry* **2013**, 56, 15-30.
18
19 33. Burkhard, P., Taylor, P., Walkinshaw, M.D. X-ray structures of small ligand-FKBP
20 complexes provide an estimate for hydrophobic interaction energies. *J. Mol. Biol.*
21 **2000**, 295(4), 953-962.
22
23 34. Cheng, L.; Pettersen, D.; Ohlsson, B.; Schell, P.; Karle, M.; Evertsson, E.; Pahlén, S.;
24 Jonforsen, M.; Plowright, A.T.; Boström, J.; Fex, T.; Thelin, A.; Hilgendorf, C.; Xue, Y.;
25 Wahlund, G.; Lindberg, W; Larsson, L.; Gustafsson, D. Discovery of the Fibrinolysis
26 Inhibitor AZD6564, Acting via Interference of a Protein–Protein Interaction **2014**, 5 (5),
27 538-543.
28
29 35. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.;
30 Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000**, 28,
31 235-242.
32
33 36. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W.,
34 Protein and ligand preparation: parameters, protocols, and influence on virtual
35 screening enrichments. *Journal of Computer-Aided Molecular Design* **2013**, 27, 221-
36 234.
37
38 37. *Schrödinger Suite 2016-2. Schrödinger, LLC, New York, 2016.*
39
40 38. Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.;
41 Tang, J., Structure of the Protease Domain of Memapsin 2 (β -Secretase) Complexed
42 with Inhibitor. *Science* **2000**, 290, 150.
43
44 39. Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.;
45 Friesner, R. A., A hierarchical approach to all-atom protein loop prediction. *Proteins*
46 **2004**, 55, 351-67.
47
48 40. Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B., On the Role of the Crystal
49 Environment in Determining Protein Side-chain Conformations. *Journal of Molecular*
50 *Biology* **2002**, 320, 597-608.
51
52 41. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.;
53 Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone
54 Parameters from ff99SB. *Journal of Chemical Theory Computation* **2015**, 11, 3696-713.
55
56 42. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An $N \cdot \log(N)$ method for
57 Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, 98, 10089-10092.
58
59 43. Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine
60 Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent
Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, 9, 3878-88.

- 1
2
3 44. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C.,
4 Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1.
5 Generalized Born. *Journal of Chemical Theory and Computation* **2012**, *8*, 1542-1555.
6
7 45. Le Grand, S.; Götz, A. W.; Walker, R. C., SPFP: Speed without compromise—A
8 mixed precision model for GPU accelerated molecular dynamics simulations. *Computer*
9 *Physics Communications* **2013**, *184*, 374-380.
10
11 46. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the
12 cartesian equations of motion of a system with constraints: molecular dynamics of n-
13 alkanes. *Journal of Computational Physics* **1977**, *23*, 327-341.
14
15 47. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and
16 use of quantum mechanical molecular models. 76. AM1: a new general purpose
17 quantum mechanical molecular model. *Journal of the American Chemical Society* **1985**,
18 *107*, 3902-3909.
19
20 48. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development
21 and testing of a general amber force field. *Journal of Computational Chemistry* **2004**,
22 *25*, 1157-74.
23
24 49. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-
25 quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational*
26 *Chemistry* **2000**, *21*, 132-146.
27
28 50. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality
29 atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of*
30 *Computational Chemistry* **2002**, *23*, 1623-41.
31
32 51. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step
33 Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical*
34 *Theory and Computation* **2015**, *11*, 1864-74.
35
36 52. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-
37 scale conformational changes with a modified generalized born model. *Proteins:*
38 *Structure, Function & Bioinformatics* **2004**, *55*, 383-394.
39
40 53. Jain, P.; Wadhwa, P. K.; Rohilla, S.; Jadhav, H. R., Rational design, synthesis and
41 in vitro evaluation of allylidene hydrazinecarboximidamide derivatives as BACE-1
42 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2016**, *26*, 33-37.
43
44 54. Neumann, U.; Ufer, M.; Jacobson, L. H.; Rouzade-Dominguez, M.-L.; Huledal, G.;
45 Kolly, C.; Löönd, R. M.; Machauer, R.; Veenstra, S. J.; Hurth, K.; Rueeger, H.; Tintelnot-
46 Blomley, M.; Staufienbiel, M.; Shimshek, D. R.; Perrot, L.; Frieauff, W.; Dubost, V.;
47 Schiller, H.; Vogg, B.; Beltz, K.; Avrameas, A.; Kretz, S.; Pezous, N.; Rondeau, J.-M.;
48 Beckmann, N.; Hartmann, A.; Vormfelde, S.; David, O. J.; Galli, B.; Ramos, R.; Graf, A.;
49 Lopez Lopez, C., The BACE-1 inhibitor CNP520 for prevention trials in Alzheimer's
50 disease. *EMBO molecular medicine* **2018**, *10*, e9316.
51
52 55. Nakahara, K.; Fuchino, K.; Komano, K.; Asada, N.; Tadano, G.; Hasegawa, T.;
53 Yamamoto, T.; Sako, Y.; Ogawa, M.; Unemura, C.; Hosono, M.; Ito, H.; Sakaguchi, G.;
54 Ando, S.; Ohnishi, S.; Kido, Y.; Fukushima, T.; Dhuyvetter, D.; Borghys, H.; Gijzen, H. J.
55 M.; Yamano, Y.; Iso, Y.; Kusakabe, K.-i., Discovery of Potent and Centrally Active 6-
56 Substituted 5-Fluoro-1,3-dihydro-oxazine β -Secretase (BACE1) Inhibitors via Active
57 Conformation Stabilization. *Journal of Medicinal Chemistry* **2018**, *61*, 5525-5546.
58
59 56. Veenstra, S. J.; Rueeger, H.; Voegtler, M.; Lueoend, R.; Holzer, P.; Hurth, K.;
60 Tintelnot-Blomley, M.; Frederiksen, M.; Rondeau, J.-M.; Jacobson, L.; Staufienbiel, M.;

- 1
2
3 Neumann, U.; Machauer, R., Discovery of amino-1,4-oxazines as potent BACE-1
4 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2018**, *28*, 2195-2200.
- 5
6 57. Low, J. D.; Bartberger, M. D.; Chen, K.; Cheng, Y.; Fielden, M. R.; Gore, V.;
7 Hickman, D.; Liu, Q.; Allen Sickmier, E.; Vargas, H. M.; Werner, J.; White, R. D.;
8 Whittington, D. A.; Wood, S.; Minatti, A. E., Development of 2-aminoxazoline 3-
9 azaxanthene β -amyloid cleaving enzyme (BACE) inhibitors with improved selectivity
10 against Cathepsin D. *MedChemComm* **2017**, *8*, 1196-1206.
- 11
12 58. Fuchino, K.; Mitsuoka, Y.; Masui, M.; Kurose, N.; Yoshida, S.; Komano, K.;
13 Yamamoto, T.; Ogawa, M.; Unemura, C.; Hosono, M.; Ito, H.; Sakaguchi, G.; Ando, S.;
14 Ohnishi, S.; Kido, Y.; Fukushima, T.; Miyajima, H.; Hiroyama, S.; Koyabu, K.; Dhuyvetter,
15 D.; Borghys, H.; Gijzen, H. J. M.; Yamano, Y.; Iso, Y.; Kusakabe, K.-i., Rational Design of
16 Novel 1,3-Oxazine Based β -Secretase (BACE1) Inhibitors: Incorporation of a Double
17 Bond To Reduce P-gp Efflux Leading to Robust A β Reduction in the Brain. *Journal of*
18 *Medicinal Chemistry* **2018**, *61*, 5122-5137.
- 19
20 59. Johansson, P.; Kaspersson, K.; Gurrell, I. K.; Bäck, E.; Eketjäll, S.; Scott, C. W.;
21 Cebers, G.; Thorne, P.; McKenzie, M. J.; Beaton, H.; Davey, P.; Kolmodin, K.; Holenz, J.;
22 Duggan, M. E.; Budd Haeberlein, S.; Bürli, R. W., Toward β -Secretase-1 Inhibitors with
23 Improved Isoform Selectivity. *Journal of Medicinal Chemistry* **2018**, *61*, 3491-3502.
- 24
25 60. Low, J. D.; Bartberger, M. D.; Cheng, Y.; Whittington, D.; Xue, Q.; Wood, S.;
26 Allen, J. R.; Minatti, A. E., Diastereoselective synthesis of fused cyclopropyl-3-amino-
27 2,4-oxazine β -amyloid cleaving enzyme (BACE) inhibitors and their biological
28 evaluation. *Bioorganic & Medicinal Chemistry Letters* **2018**, *28*, 1111-1115.
- 29
30 61. Tanaka, Y.; Aikawa, K.; Nishida, G.; Homma, M.; Sogabe, S.; Igaki, S.; Hayano, Y.;
31 Sameshima, T.; Miyahisa, I.; Kawamoto, T.; Tawada, M.; Imai, Y.; Inazuka, M.; Cho, N.;
32 Imaeda, Y.; Ishikawa, T., Discovery of Potent Mcl-1/Bcl-xL Dual Inhibitors by Using a
33 Hybridization Strategy Based on Structural Analysis of Target Proteins. *Journal of*
34 *Medicinal Chemistry* **2013**, *56*, 9635-9645.
- 35
36 62. Stewart, M. L.; Fire, E.; Keating, A. E.; Walensky, L. D., The MCL-1 BH3 helix is an
37 exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nature Chemical Biology* **2010**, *6*,
38 595.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

