



Performance evaluation of molecular docking and free energy calculations protocols using the D3R Grand Challenge 4 dataset

Eddy Elisée¹ · Vytautas Gapsys² · Nawel Mele¹ · Ludovic Chaput^{1,3} · Edithe Selwa¹ · Bert L. de Groot² · Bogdan I. Iorga¹

Received: 16 June 2019 / Accepted: 9 October 2019 / Published online: 1 November 2019
© Springer Nature Switzerland AG 2019

Abstract

Using the D3R Grand Challenge 4 dataset containing Beta-secretase 1 (BACE) and Cathepsin S (CatS) inhibitors, we have evaluated the performance of our in-house docking workflow that involves in the first step the selection of the most suitable docking software for the system of interest based on structural and functional information available in public databases, followed by the docking of the dataset to predict the binding modes and ranking of ligands. The macrocyclic nature of the BACE ligands brought additional challenges, which were dealt with by a careful preparation of the three-dimensional input structures for ligands. This provided top-performing predictions for BACE, in contrast with CatS, where the predictions in the absence of guiding constraints provided poor results. These results highlight the importance of previous structural knowledge that is needed for correct predictions on some challenging targets. After the end of the challenge, we also carried out free energy calculations (i.e. in a non-blinded manner) for CatS using the pmx software and several force fields (AMBER, Charmm). Using knowledge-based starting pose construction allowed reaching remarkable accuracy for the CatS free energy estimates. Interestingly, we show that the use of a consensus result, by averaging the results from different force fields, increases the prediction accuracy.

Keywords Molecular docking · Free energy calculations · Molecular dynamics · Pmx · D3R challenge · Beta secretase 1 · Cathepsin S · Inhibitors

Research reported in this publication was supported by grants ANR-10-LABX-33 (LabEx LERMIT) and ANR-14-JAMR-0002 (JPIAMR) from the French National Research Agency (ANR), by the Région Ile-de-France (DIM Malinf), by the Université Paris-Saclay (Globetalkers 2019) and by European Research Council grant ERC-2012-ADG_20120314 (Grant Agreement 322947).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10822-019-00232-w>) contains supplementary material, which is available to authorized users.

✉ Bert L. de Groot
bgroot@gwdg.de

✉ Bogdan I. Iorga
bogdan.iorga@cnrs.fr

¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, Labex LERMIT, Gif-sur-Yvette, France

² Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

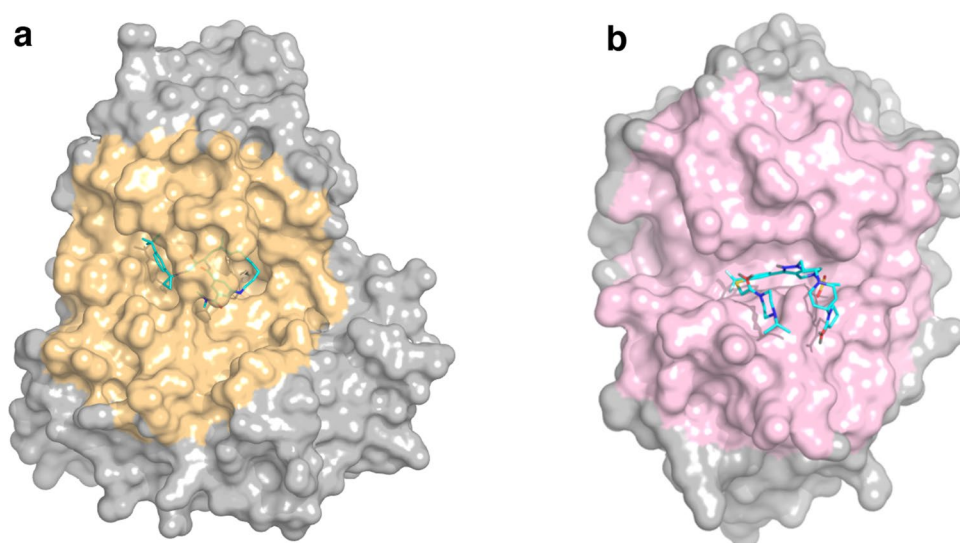
³ Sorbonne Université, UPMC Paris 06, Institut National de la Santé et de la Recherche Médicale, Unité Mixte de Recherche S 1155, Paris, France

Introduction

Over the years, the blinded prediction challenges regularly organized by the Drug Design Data Resource (D3R, <https://drugdesigndata.org/>) became valuable opportunities to assess the quality and performance of in-house methodologies often used in computer-aided drug discovery projects. The concept is simple: datasets are generously provided by known pharmaceutical companies and presented to the computational chemistry community, which is invited to blindly predict binding modes and affinities of ligands.

The D3R Grand Challenge 4 was organized in 2018 and was based on two protein targets: cathepsin S (CatS, Fig. 1b), which was already present in the previous D3R Grand Challenge 3, and beta-secretase 1 (BACE, Fig. 1a). Cathepsin S is a lysosomal cysteine protease involved, inter alia, in antigen presentation, nociception, itch, pain and whose regulation may be important in several diseases as psoriasis, rheumatoid arthritis and glioblastoma [1–9]. Beta-secretase 1 is an aspartic acid protease involved in the formation of myelin

Fig. 1 Solid surface representation of representative crystal structures of BACE (**a**, PDB code 3K5C [16]) and cathepsin S (**b**, PDB code 1GLO [17]) with the ligands (**BACE_68** and **CatS_86**, respectively) in stick representation and colored in cyan. The binding sites, defined as spheres of 20 Å around the atom OG1 of THR72 in (**a**) and around a virtual point with coordinates $[-7, 8, -4]$ in (**b**), are colored in pink and orange, respectively



sheaths in peripheral nerve cells and in Alzheimer's disease by cleavage of the amyloid precursor protein, leading, in the end, to amyloid- β peptide aggregation [10–13]. However, recent publications relate the possible fail of BACE inhibitors in Alzheimer's trials [14, 15].

The BACE subchallenge included three stages. In Phase 1A, the participants were asked to predict the crystallographic poses of 20 BACE ligands (Fig. 2), the affinity ranking for 154 BACE ligands (Fig. S1 in the Electronic Supplementary Material) and the binding affinities for a designated free energy subset of 34 BACE ligands (Fig. 3). In Phase 1B, organizers released the corresponding

receptor structures (without ligands) for the 20 BACE ligands composing the pose prediction subset, and the participants could repeat the pose prediction using this additional information. In Phase 2, organizers released the complexes from the pose prediction subset and the participants were required to repeat the affinity predictions for the subset of 154 ligands and the relative free energy for the subset of 34 compounds as in Phase 1A, but taking into account the structural information from the 20 released protein-ligand complexes. The CatS subchallenge included a single stage, Phase 2, with ranking prediction for two datasets of 459 ligands (structures not shown) and 39 ligands (Fig. 4), the latter being designed for relative free energy calculations.

Fig. 2 Chemical structures of the D3R GC4 BACE pose prediction dataset

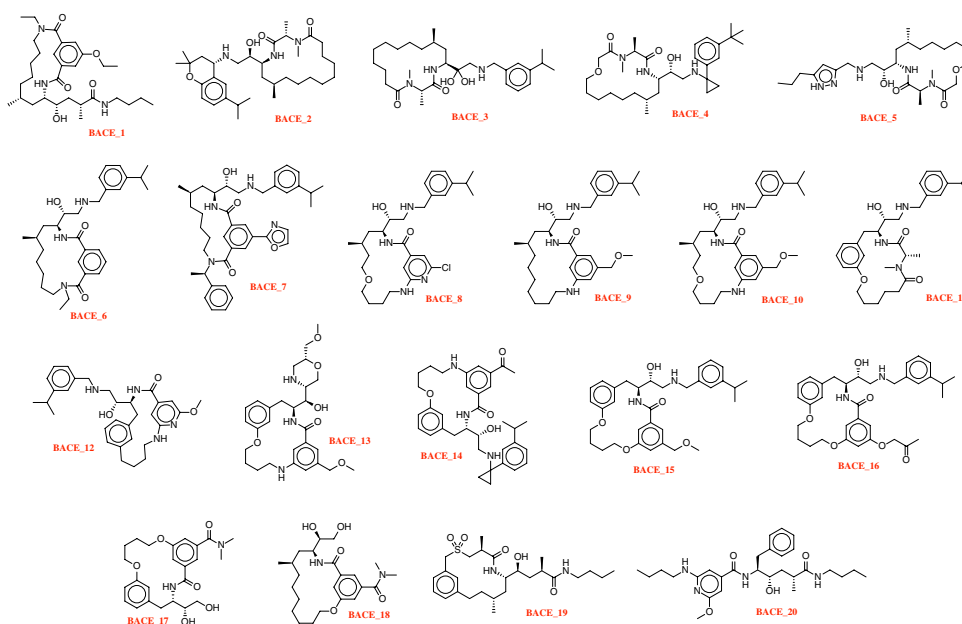
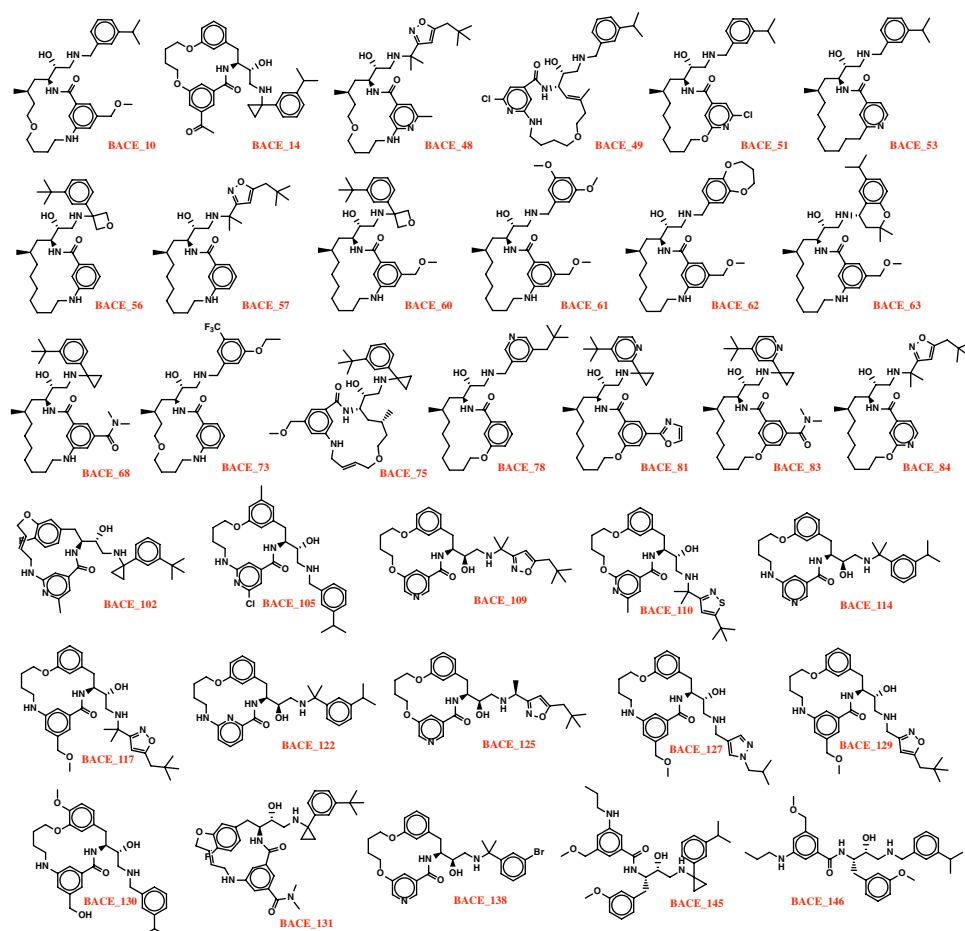


Fig. 3 Chemical structures of the D3R GC4 BACE free energy dataset



Methods

Protein structures

Analysis of Protein Data Bank [18] revealed 341 available crystal structures of human BACE but only 337 of them, including four synthetic structures (1M4H [19], 4TRW [20], 4TRY [20], 4TRZ [20]), were considered in this work since two structures were not released when the panel of proteins has been built (6EQM [21] and 6DMI [22]) and two proteins contained useless extra-domains (1UJJ [23], 1UJK [23]). Furthermore, four apo structures were found but only one was complete, namely without missing residues (1SGZ [24]).

The three-dimensional alignment of all these structures revealed a perfectly conserved fold except for three regions: the first turn near the active site (VAL309-ASP318), the second turn (LYS9-GLY13) with open (e.g. structure 2VIZ [25]) and closed (e.g. structure 2VIJ [26]) conformations, and the third turn (PRO70-GLY74).

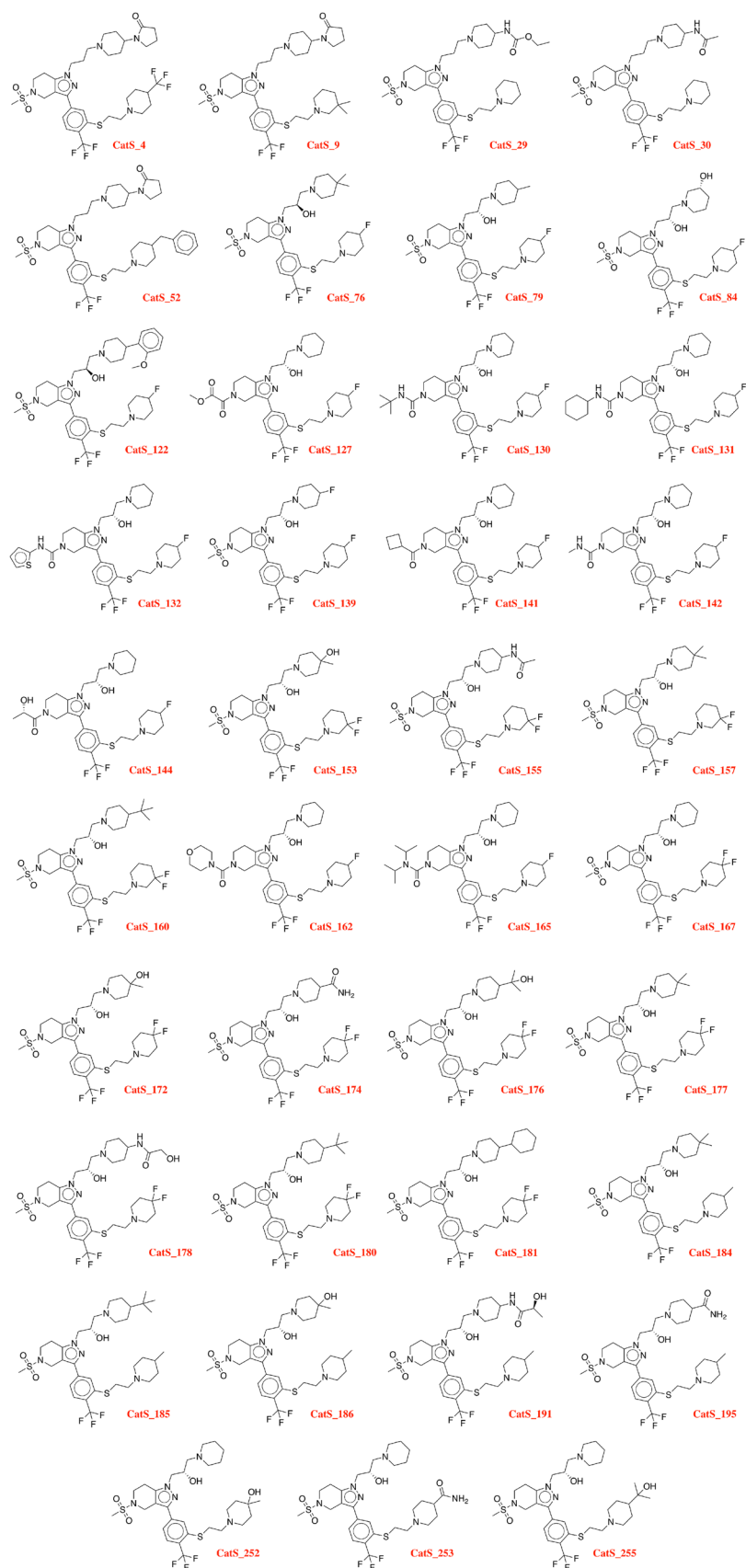
The proteins were protonated at physiological pH using the appropriate tool corresponding to each docking software: Hermes for GOLD and Maestro for Glide and Vina.

The structural analysis for cathepsin S was carried out during our participation to the D3R Grand Challenge 3 in 2017 [27] and, given the similarity of the CatS ligands proposed in the previous and the present challenges, we decided to use in this study the same protocol, with the exception of the constraints during the docking process, which were not used anymore.

Ligands

Three-dimensional coordinates for all compounds were generated in MOL2 format using two protocols: (i) random conformations were built from SMILES strings using LIGPREP v45011 (Schrödinger, <http://www.schrodinger.com>), software also used to adjust protonation states at physiological pH; and (ii) bioactive conformations were built manually from the closest macrocycle-containing BACE ligands available in the PDB (see the Electronic Supplementary Material for the correspondence between these structures) or from the closest CatS ligands released during the previous D3R GC3 challenge (haan-CatS_11, wcgq-CatS_10 and mekm-CatS_13), using UCSF Chimera [28].

Fig. 4 Chemical structures of the D3R GC4 CatS free energy dataset



The adjustment of protonation states at physiological pH for all ligands makes that the CatS ligands have a total charge of + 2, which is in agreement with the instructions provided by the organizers of D3R GC4 challenge stating that “all compounds in both the BACE and CatS free energy sets [are expected to have] a charge of + 2 at the assay pH values of 4.5 (BACE) and 5.0 (CatS).” In contrast, our BACE ligands have a total charge of + 1, which is in apparent contradiction with these instructions. We were concerned about this difference and we investigated in more detail the BACE binding site residues that would interact with this second charge (which in individual BACE ligands should be positioned in different regions of space). We could evidence no specific ionic interaction and therefore, we continued our study with ligands having a total charge of + 1. To the best of our knowledge, there is no specific study providing an experimental proof that the BACE inhibitors from this series have a total charge of + 2 when bound to the protein.

We organized the BACE ligands in four groups, according to the nature of the macrocyclic core: (A) aliphatic macrocycle, (B) one aromatic ring, (C) two aromatic rings, (D) no macrocycle. The substituents bound to the macrocycle are hereinafter referred to as “lateral chains”. Table S1 in the Electronic Supplementary Material sums up the design plan of the 158 ligands.

Preliminary docking

A training set was created with 14 BACE structures containing ligands similar with those from the D3R GC4 dataset (similarity evaluated using Tanimoto method). IC50 values could be retrieved from BindingDB database [29] for 12 of them (see Table S2 in the Electronic Supplementary Material).

The alignment of these structures showed a reasonable conserved three-dimensional structure, except for two amino acids: (a) the conformation of ARG235 in the structure 3DV1 [30] clashes with the ligands from structures 2VIJ [26], 2VJ7 [31], 2VNM [32], 3K5C [16] and 4DPF [33], whereas other equivalent arginines, e.g. ARG296 in the structure 2VIJ [26], show no clashes; (b) the conformation of GLN73 in the structure 3DV1 [30], like most equivalent glutamines in this position, clashes with all ligands except those from the 3K5C [16], 3K5F [34] and 3IVH [35] structures.

We finally chose the 3K5C structure for all subsequent docking studies as its active site is open enough to accept all other ligands from the training set (only 4DPF ligand has a phenyl substituent, as in **BACE_69** and **BACE_82**, which is slightly clashing with the 3K5C surface, but no amino acid in its environment seems to be an issue).

In the next step, several docking software and scoring functions have been tested for their ability to predict relative

affinity ranking and reproduce the protein-ligand complexes from this training dataset: GOLD [36] with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina [37] and Glide (Schrödinger, <http://www.schrodinger.com>). Default parameters were used in all cases for docking, except with GOLD, where a search efficiency of 200% and the “flip ring corners” option were used in order to better explore the conformational space of the ligands macrocycles. For Glide, the standard precision (SP) mode was used. The binding sites were considered with GOLD as spheres with a 20 Å radius around the OG1 atom of THR72 (numbering from the 3K5C structure). With Vina, the binding sites were defined as a $62 \times 64 \times 80 \text{ \AA}^3$ cube centered on the same atom. The protein structure (3K5C) was considered to be rigid during the docking process and the ligands fully flexible. Fifty docking poses were generated for each ligand from the training set.

Phase 1A docking

From the preliminary analysis, GOLD with the GoldScore scoring function were identified as the most adapted for the BACE system (see “Preliminary analysis” section and Table 1 for a more detailed discussion), and therefore were used for Phase 1 predictions with default values except search efficiency which was set to 200%. To explore ring conformations and the ligand flexibility, the options “flip ring corners”, “flip pyramidal N” and “flip amide bonds” were activated. All the ligands were docked onto the 3K5C [16] structure, the ligands being fully flexible and the receptor rigid. The docking calculations were carried out with two hydrogen bond constraints involving the backbone oxygen atoms of residues GLY230 and GLY34 and any atoms of the

Table 1 Preliminary evaluation of the training dataset

| Software/scoring function | Method for generating ligand 3D structure | Average RMSD for best pose (Å) |
|---------------------------|---|--------------------------------|
| Glide/SP | SMILES | 3.24 |
| GOLD/GoldScore | SMILES | 2.48 |
| GOLD/ChemScore | SMILES | 3.14 |
| GOLD/ChemPLP | SMILES | 2.95 |
| GOLD/ASP | SMILES | 3.50 |
| Vina | SMILES | 6.87 |
| Glide/SP | PDB | 3.14 |
| GOLD/GoldScore | PDB | 2.35 |
| GOLD/ChemScore | PDB | 2.85 |
| GOLD/ChemPLP | PDB | 3.27 |
| GOLD/ASP | PDB | 3.19 |
| Vina | PDB | 6.41 |

See Table S2 for the list of ligands with known structures composing the training dataset

ligands. Additional docking calculations were also carried out with ligands generated from SMILES, and also without any constraints. In all these cases, as expected, the poses are slightly worse compared to those obtained with constraints, and only the latter were submitted.

Analysis of the training dataset has emphasized a conserved hydrogen bond in all the ligands involving the backbone oxygen atom of residue GLY230 and also one hydrogen bond, found in most of the ligands, involving the backbone oxygen atom of residue GLY34. Therefore, docking simulations were carried out with hydrogen bond constraints on either or both oxygen atoms, and also without constraints, for all compounds from the BACE D3RGC4 dataset, with 100 docking poses generated for each ligand.

The RMSD between the common substructure of the macrocyclic core and the side chain parts of the ligand and of one of PDB ligands AR9 (3DV1 [30]), OBI (3K5C [16]), OLG (4DPF [33]), OXA (4GMI), 1R8 (4KE0 [38]) and BAV (3DV5 [30]) were calculated using an in house developed script based on CACTVS Chemoinformatics Toolkit (<http://www.xemistry.com/>). The two RMSD values were added and the poses with the smaller sum were selected for submission (1 score per compound).

The second submission (1 score per compound) consisted of the best ranking pose for each docking.

Protein structures were converted into PDB format for submission using UCSF Chimera [28], and the docking poses were converted into MOL format using CORINA version 3.60 (<http://www.molecular-networks.com>) (the MOL format corresponds to the SDF output format in CORINA).

Phase 1B docking

After the release of the 20 crystal structures corresponding to the compounds **BACE_1** to **BACE_20** (but without the coordinates of the ligands), we carried out docking calculations with each ligand individually onto the corresponding structure released at the end of Phase 1A (GOLD with the GoldScore scoring function, 100 docking poses generated for each ligand). Each ligand was individually docked onto the corresponding structure released at the end of Phase 1A. The docking calculation was carried out with two hydrogen bond constraints involving the backbone oxygen atoms of residues GLY230 and GLY34 and any atoms of the ligands.

Phase 2 docking

As no significantly new structural information was brought by the 20 protein-ligand complexes released at the end of Phase 1B, for the BACE subchallenge in Phase 2 we have submitted the same files as in Phase 1A.

For the CatS subchallenge we used our protocol from Grand Challenge 3 (2017) [27] and performed the docking

with GOLD using ASP scoring function and 100 poses per ligand. In contrast with our previous predictions of CatS inhibitors [27], this time no constraints were used during the docking process.

Free energy calculations

All free energy calculations were carried out after the end of the challenge (i.e. in a non-blinded manner).

The D3R GC4 cathepsin S free energy dataset was composed of 39 structurally similar CatS ligands (Fig. 4).

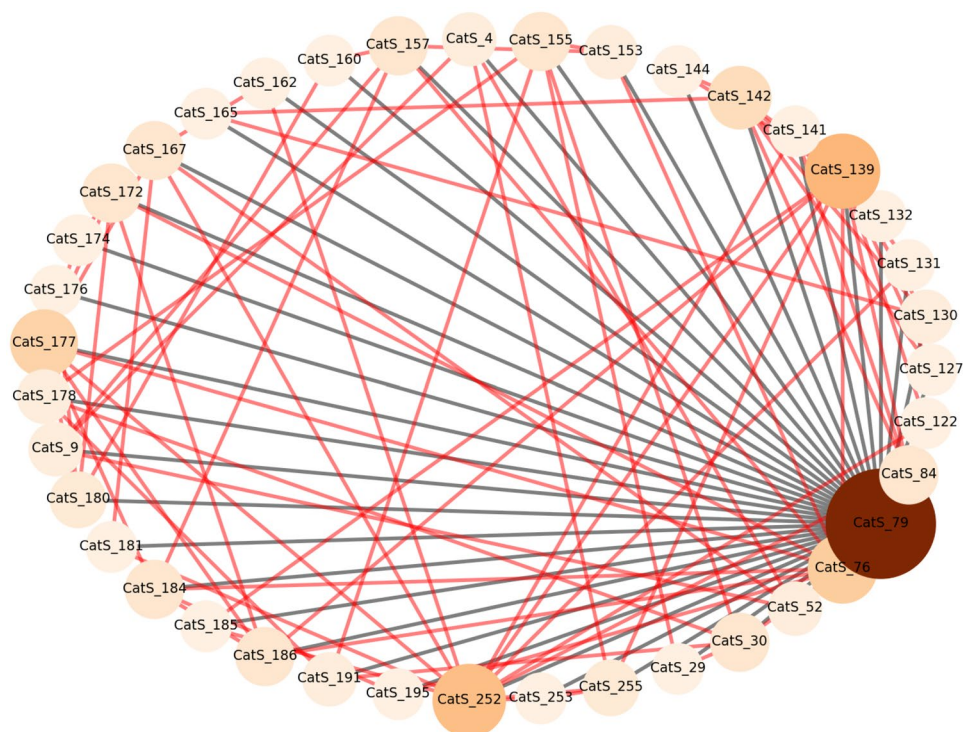
Free energies for Cathepsin S binders were estimated using non-equilibrium free energy calculation protocol [39]. Firstly, a mapping between ligands was established (Fig. 5): ligand **CatS_79** was used as an initial reference connecting to the other molecules (gray edges in the graph). This way double free energy differences ($\Delta\Delta G$) for every compound with respect to **CatS_79** could be evaluated. Subsequently, redundancies in the graph were introduced (red edges) to allow for formation of cycles that could be further used to correct for under-sampling [40]. In total 109 transitions were considered.

For the free energy calculations of CatS inhibitors we used the same ligand structures that were used as input for docking calculations (see "Ligands" section for details).

For every transition between two ligands, hybrid structures and topologies were generated using pmx package [41]. Prior to starting molecular dynamics simulations, the solvated systems were energy minimized by keeping the position restraints of $1000 \text{ kJ}\cdot\text{mol}^{-1}\text{nm}^{-2}$ on all heavy atoms. Afterwards, 10 ps constant temperature and volume simulations were performed still retaining the position restraints on heavy atoms. Starting from the end states of the NVT simulations, equilibrium simulations in isothermal-isobaric conditions for the ligands in solution and bound to protein were performed for 6 ns. Subsequently, 80 snapshots were extracted equidistantly in time from the last 4 ns of each trajectory and rapid 200 ps transitions were spawned driving the system from one physical state (ligand A) to another (ligand B). The non-equilibrium work values were recorded for every transition and related to the free energy difference based on the Crooks Fluctuation Theorem [42]. Maximum likelihood estimator was used to estimate free energy difference [43]. The whole procedure, including equilibrium sampling and non-equilibrium transitions, was repeated three times. The final free energy estimate was calculated as a mean of the three replicas. The errors for individual free energy differences were calculated as standard errors of the three repeats.

All the molecular dynamics simulations were performed with Gromacs [44]. Two protein force fields were used: Amber99SB*ILDN [45–47] and Charmm36m [48]. For ligands the Generalized Amber Force Field (GAFF

Fig. 5 Graph of CatS ligand mappings for relative free energy calculations. Circle size and color encode the node connectivity, i.e. the number of edges connected to a node (ligand). Gray edges mark the connections between every ligand and compound **CatS_79**; the additional red edges have been introduced as redundancies to allow for cycle closure correction



v2.1) [49] was used in combination with Amber99sb*ILDN. Atom types and charges for ligands in Charmm force field were assigned using MATCH algorithm [50]; CGenFF 4.1 [51] was used for bonded parameters. The systems were solvated with TIP3P water [52]. Na^+ and Cl^- ions were added to neutralize the simulation box and reach 150 mM salt concentration. Particle Mesh Ewald [53, 54] was used to treat long range electrostatics with the real space cutoff of 1.1 nm, Fourier spacing of 0.12 nm. Van der Waals interactions were smoothly switched off between 1.0 and 1.1 nm. All bonds were constrained by the LINCS [55] algorithm. Temperature was kept at 298 K by means of the velocity rescaling thermostat [56] with the time constant of 0.1 ps. Pressure of 1 bar was controlled with the Parrinello-Rahman barostat [57] with the time constant of 5 ps.

For reconstructing ΔG values from double free energy differences ($\Delta\Delta G$) **CatS_79** was used as a reference. Free energies were calculated with GAFF (Amber) and CGenFF (Charmm) force fields. Furthermore, a consensus result was obtained by averaging the results from GAFF and CGenFF force fields [58]. The overall calculation accuracy was assessed by means of root mean squared error (RMSE), Pearson's, Kendall's and Spearman's correlation coefficients. The errors for these measures were obtained by means of bootstrapping. The bootstrap procedure followed closely the one described in previous D3R challenges [59]: namely, 10,000 samplings with replacement were performed from the calculated and experimental value sets. The experimental values were modified by adding Gaussian noise with

the mean of zero and standard deviation of $\text{RT} \ln I_{err}$, with $I_{err} = 2.5$.

The D3R GC4 BACE free energy dataset was composed of 34 structurally similar BACE ligands (Fig. 3). During the free energy calculations we have encountered difficulties with the stability of simulations and convergence problems. These might be due, at least in part, to the variations in the macrocyclic scaffold of BACE ligands from this dataset.

Graphics

The figures representing protein-ligand complexes were generated with Pymol [60], and the chemical structures were drawn using ChemDraw Professional 16.0 (PerkinElmer Informatics). The CatS ligand perturbation network was built with NetworkX [61]. The results of CatS calculations were visualized using Matplotlib [62] library.

Results and discussion

From our participation to previous docking and virtual screening challenges, SAMPL3 (2011) [63], SAMPL4 (2013) [64], CSAR (2014) [65], D3R Grand Challenge (2015) [66], D3R Grand Challenge 2 (2016) [67] and D3R Grand Challenge 3 (2017) [27], we developed a protocol for docking and virtual screening that proved to be relatively successful. This protocol involves two steps, the first one representing a preliminary analysis of publicly

available structural and biochemical data in order to identify the most appropriate docking software and scoring function for the system of interest. In the second step, we use these parameters for the prediction of binding modes (pose prediction) and of relative affinities of ligands (scoring). Enhanced genetic algorithm parameters for docking (a search efficiency of 200%) proved to be beneficial in our previous studies [27, 63–67], especially for big and flexible ligands. Thus, in this work we used the same parameters in order to ensure an adequate conformational sampling of docking conformations.

Preliminary analysis

In the preliminary analysis step, the ligands from the training set were docked using the structure 3K5C [16] and different combinations of docking software and scoring functions, and the results were analyzed for accuracy in terms of pose prediction and ranking. In order to evaluate the accuracy of docking and scoring, we have considered the lowest RMSD value and the RMSD value of the best ranking pose for each combination protein-ligand-(docking software)-(scoring function) reported in Table 1. The RMSD values were computed for all atoms, with the crystallographic structures of these ligands as reference (see Table S2 for the list of ligands with known structures composing the training dataset.) Table 1 shows that GOLD with the GoldScore scoring function clearly outperformed the other docking programs and scoring functions in the pose prediction of the training set. Vina provided very poor results, with most of the docking conformations positioned outside the binding site, whereas GOLD/GoldScore, followed by GOLD/ChemScore and Glide/SP, could reproduce rather well the native protein-ligand complexes. In these conditions GOLD with GoldScore with the 3K5C structure were used for the following steps.

No significant correlation could be found between the docking score and biological activity (pIC50) for the 12 compounds from the training set for which experimental biological data were available.

Phase 1A: pose prediction and ranking (BACE)

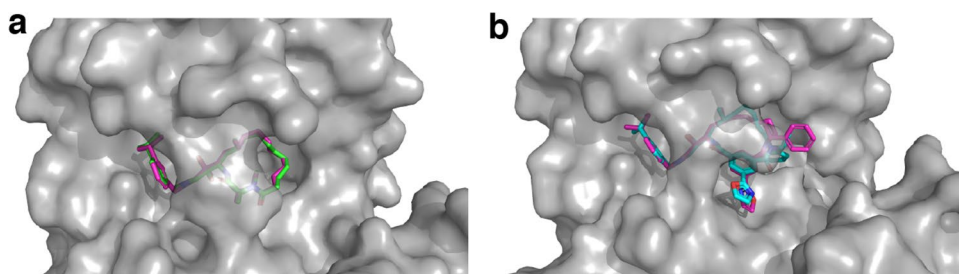
The 154 BACE ligands were docked on 3K5C [16] structure using GOLD and the GoldScore scoring function to generate 50 poses for each ligand. Three versions of the protocol were tested (see the "Methods" section for details) with and without hydrogen bonds constraints. For the ligands belonging to a group for which crystal structures were available, the RMSD was calculated between the common substructure of the macrocyclic core and the side chain parts of the ligand and of one of PDB ligands AR9 (3DV1 [30]), OBI (3K5C [16]), OLG (4DPF [33]), OXA (4GMI), 1R8 (4KE0 [38]) and BAV (3DV5 [30]) using an in house developed script based on CACTVS Chemoinformatics Toolkit (<http://www.xemistry.com/>). The two RMSD values were added and the poses with the smaller sum were selected for submission in phases 1 and 2. RMSD values compared with the part of the native ligands from the crystallographic structures were calculated for all docking poses. We obtained RMSD values calculated on the common substructure with ligands with known structures ranging from 1.52 to 1.90 Å for the "Best RMSD" prediction and from 1.81 to 2.30 Å for the "Best Score" prediction. The results demonstrate that the use of constraints induced similar performances in the two cases and, as expected, the poses without constraints are slightly worse compared to those obtained with constraints, and only the latter were submitted.

The release of the Phase 1A results showed that our two BACE ranking predictions performed the best in this competition, and that the "Best RMSD" pose prediction was ranked 7th. The pose predictions for the best and the worst compounds are presented in Fig. 6.

Phase 1B: pose prediction (BACE)

In Phase 1B, we carried out docking calculations using the 20 crystal structures corresponding to the compounds **BACE_1** to **BACE_20** that were released at the end of Phase 1A (but without the coordinates of the ligands) with their native ligands. The re-docking of ligands into their native structures provided only a slight improvement in the pose prediction. Similar to the phase 1A, adding constraints

Fig. 6 Comparison of our BACE Phase 1A best and worst pose predictions for compounds **BACE_3** (a, green) and **BACE_7** (b, cyan) with the corresponding BACE complexes that were released at the end of the challenge (RMSD 0.39 Å and 2.60 Å, respectively)



did not improve the predictions, so only simulations with one constraint were submitted.

Phase 2: ranking (BACE and CatS)

As no significantly new structural information was brought by the 20 protein-ligand complexes released at the end of Phase 1B, for the BACE subchallenge in Phase 2 we decided to submit the same results as in Phase 1A.

Concerning CatS simulations, similarly as it was done for BACE, RMSD calculations between the common substructure of the ligand and of one of PDB ligands B8J (5QBU), BCJ (5QC6), BJJ (5QCG), B9S (5QC1) and BFV (5QC8) were calculated using an in house developed script based on CACTVS Cheminformatics Toolkit (<http://www.xemistry.com/>). The ranking with the best final RMSD poses were selected for submission (see "Phase 1A docking" section for details about RMSD calculation).

The pose prediction is variable within the CatS docking simulations, inducing poorer results compared with BACE. Some ligands are incorrectly positioned as illustrated in Fig. 7a, but when the core of the ligand is correctly positioned (Fig. 7b), a significant improvement in RMSD is observed.

Free energy calculations

For timing reasons, no free energy predictions were submitted during the challenge, and all free energy calculations are presented retrospectively.

Overall, calculated free energies are well within 1 kcal/mol deviation from the experimental measurements in terms of AUE and $RMSE_C$ (Table 2, Fig. 8). In terms of absolute agreement with the experimental ΔG values, as measured by AUE and $RMSE_C$, our calculations using GAFF force field outperformed all the other submissions to the D3R GC4 (Fig. 8). When predicting the overall trend (Pearson correlation) and ligand ranking by their binding affinity (Kendall's and Spearman's correlation), GAFF based calculations were only superseded by another set of submissions utilizing the same ligand force field in an alchemical free energy calculation setup. The free energy estimates based on the CGenFF and Charmm36m force field combination performed slightly, but consistently worse with respect to all the considered quality measures.

Prediction accuracy further increased when applying the consensus approach. This observation is consistent with the previously reported observations for protein thermostabilities [58], DNA nucleotide mutations [68], protein-ligand interactions [69].

Of the 39 estimated free energy values, the consensus force field approach only in 3 cases shows a deviation from experiment larger than 1 kcal/mol (Tables S3 and S4). One of the major sources of error in the current calculation procedure comes from the starting pose definition. In the current setup, pose construction based on the structural data from the previous D3R GC3 challenge (see "Ligands" section) proved to yield highly accurate results. In other attempts, where we probed the accuracy of alchemical predictions when starting from a distorted set of poses the estimation accuracy markedly decreased. All in all, starting

Fig. 7 Comparison of our CatS Phase 2 best ranking poses for compounds **CatS_105** (a, cyan) and **CatS_245** (b, green) with the ligand BCJ (magenta) from the CatS crystal structure 5QC6 (RMSD 11.20 Å and 1.88 Å, respectively, on the macrocyclic core)

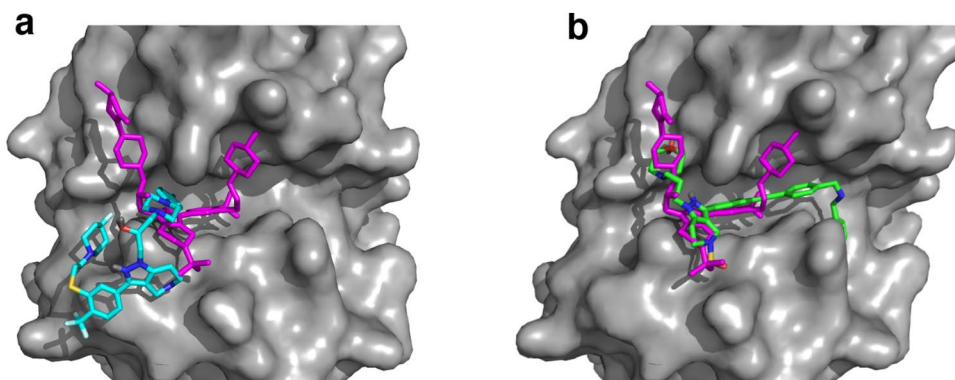


Table 2 Summary of the calculated alchemical free energy accuracies in terms of average unsigned error (AUE), root mean squared error (RMSE), Pearson's, Kendall's and Spearman's correlation coefficients

| Force field | AUE (kcal/mol) | $RMSE_C$ (kcal/mol) | Pearson's r | Kendall's τ | Spearman's ρ |
|-------------|-----------------|---------------------|-----------------|------------------|-------------------|
| GAFF | 0.40 ± 0.09 | 0.45 ± 0.10 | 0.72 ± 0.14 | 0.49 ± 0.12 | 0.67 ± 0.16 |
| CGenFF | 0.48 ± 0.10 | 0.50 ± 0.11 | 0.60 ± 0.16 | 0.39 ± 0.13 | 0.54 ± 0.17 |
| Consensus | 0.37 ± 0.10 | 0.43 ± 0.10 | 0.72 ± 0.15 | 0.49 ± 0.12 | 0.67 ± 0.16 |

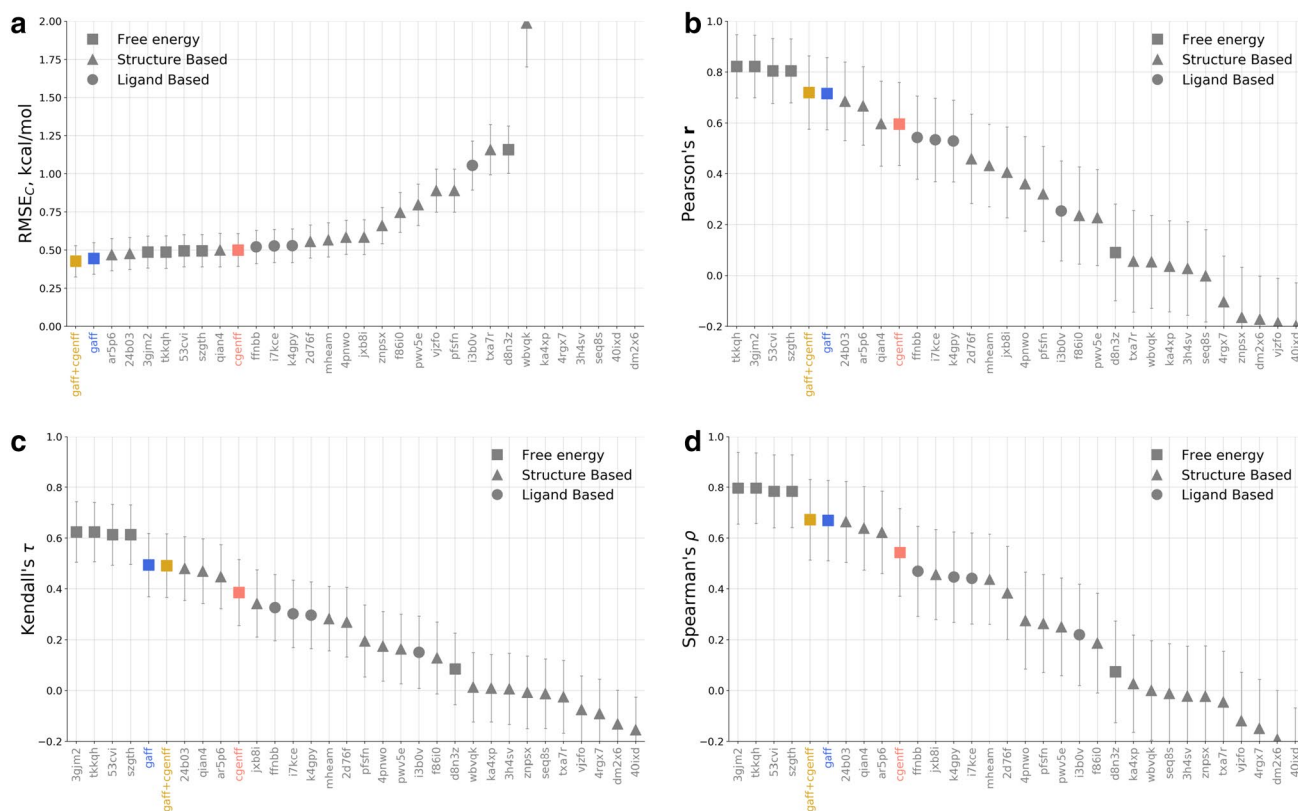


Fig. 8 Summary of the CatS free energy calculation results. The calculations performed in this work (GAFF—blue, CGenFF—red, consensus—yellow) are shown together with the other submissions: **a** RMSE_C, **b** Pearson correlation **c**, Kendall's τ and **d** Spearman's correlation

pose generation needs to be taken with care as it largely modulates accuracy achievable with alchemical free energy calculations.

Conclusions

Using the D3R Grand Challenge 4 dataset containing beta-secretase 1 (BACE) and cathepsin S (CatS) inhibitors, we have evaluated the performance of our in-house docking workflow that involves in the first step the selection of the most suitable docking software for the system of interest based on structural and functional information available in public databases, followed by the docking of the dataset to predict the binding modes and ranking of ligands. The macrocyclic nature of the BACE ligands brought additional challenges, which were dealt with by a careful preparation of the three-dimensional input structures for ligands. This provided top-performing predictions for BACE, in contrast with CatS, where the predictions in the absence of guiding constraints provided poor results. These results highlight the importance of previous structural knowledge that is needed for correct predictions on some challenging targets. After the end of the challenge, we also carried out free energy

calculations (i.e. in a non-blinded manner) for CatS using the pmx software and several force fields (AMBER, Charmm). Using knowledge-based starting pose construction allowed reaching remarkable accuracy for the CatS free energy estimates. Interestingly, we show that the use of a consensus result, by averaging the results from different force fields, increases the prediction accuracy.

Funding The funding was provided by Agence Nationale de la Recherche (FR) (Grant Nos. ANR-10-LABX-33, ANR-14-JAMR-0002); Conseil Régional, Île-de-France (DIM Malinf); Université Paris-Saclay (Globetalkers 2019); H2020 European Research Council (ERC-2012-ADG_20120314, Grant Agreement 322947).

References

- Reddy VB, Sun S, Azimi E, Elmariam SB, Dong X, Lerner EA (2015) Redefining the concept of protease-activated receptors: cathepsin S evokes itch via activation of Mrgprs. *Nat Commun* 6:7864. <https://doi.org/10.1038/ncomms8864>
- Ainscough JS, Macleod T, McGonagle D, Brakefield R, Baron JM, Alase A, Wittmann M, Stacey M (2017) Cathepsin S is the major activator of the psoriasis-associated proinflammatory

- cytokine IL-36 γ . *Proc Natl Acad Sci USA* 114(13):E2748–E2757. <https://doi.org/10.1073/pnas.1620954114>
3. Elmariah SB, Reddy VB, Lerner EA (2014) Cathepsin S signals via PAR2 and generates a novel tethered ligand receptor agonist. *PLoS ONE* 9(6):e99702. <https://doi.org/10.1371/journal.pone.0099702>
 4. Xu J, Wang H, Ding K, Lu X, Li T, Wang J, Wang C (2013) Inhibition of cathepsin S produces neuroprotective effects after traumatic brain injury in mice. *Mediat Inflamm* 187:873. <https://doi.org/10.1155/2013/187873>
 5. Thurmond RL, Sun S, Karlsson L, Edwards JP (2005) Cathepsin S inhibitors as novel immunomodulators. *Curr Opin Investig Drugs* 6(5):473–482
 6. Link JO, Zipfel S (2006) Advances in cathepsin S inhibitor design. *Curr Opin Drug Discov Devel* 9(4):471–482
 7. Wiener JJM, Sun S, Thurmond RL (2010) Recent advances in the design of cathepsin S inhibitors. *Curr Top Med Chem* 10(7):717–732. <https://doi.org/10.2174/156802610791113432>
 8. Lee-Dutra A, Wiener DK, Sun S (2011) Cathepsin S inhibitors: 2004–2010. *Expert Opin Ther Pat* 21(3):311–337. <https://doi.org/10.1517/13543776.2011.553800>
 9. Wilkinson RDA, Williams R, Scott CJ, Burden RE (2015) Cathepsin S: therapeutic, diagnostic, and prognostic potential. *Biol Chem* 396(8):867–882. <https://doi.org/10.1515/hsz-2015-0114>
 10. Vassar R, Bennett BD, Babu-Khan S, Kahn S, Mendiaz EA, Denis P, Teplow DB, Ross S, Amarante P, Loeloff R, Luo Y, Fisher S, Fuller J, Edenson S, Lile J, Jarosinski MA, Biere AL, Curran E, Burgess T, Louis JC, Collins F, Treanor J, Rogers G, Citron M (1999) Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* 286(5440):735–741. <https://doi.org/10.1126/science.286.5440.735>
 11. Willem M, Garratt AN, Novak B, Citron M, Kaufmann S, Rittger A, DeStrooper B, Saftig P, Birchmeier C, Haass C (2006) Control of peripheral nerve myelination by the beta-secretase BACE1. *Science* 314(5799):664–666. <https://doi.org/10.1126/science.1132341>
 12. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, Hoyte K, Gustafson A, Liu Y, Lu Y, Bhangale T, Graham RR, Huttenlocher J, Bjornsdottir G, Andreassen OA, Jönsson EG, Palotie A, Behrens TW, Magnusson OT, Kong A, Thorsteinsdottir U, Watts RJ, Stefansson K (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488(7409):96–99. <https://doi.org/10.1038/nature11283>
 13. Baxter EW, Conway KA, Kennis L, Bischoff F, Mercken MH, Winter HLD, Reynolds CH, Tounge BA, Luo C, Scott MK, Huang Y, Braeken M, Pieters SMA, Berthelot DJC, Masure S, Bruinzeel WD, Jordan AD, Parker MH, Boyd RE, Qu J, Alexander RS, Brenneman DE, Reitz AB (2007) 2-Amino-3,4-dihydroquinazolines as inhibitors of BACE-1 (beta-site APP cleaving enzyme): use of structure based design to convert a micromolar hit into a nanomolar lead. *J Med Chem* 50(18):4261–4264. <https://doi.org/10.1021/jm0705408>
 14. Panza F, Lozupone M, Solfrizzi V, Sardone R, Piccininni C, Dibello V, Stallone R, Giannelli G, Bellomo A, Greco A, Daniele A, Seripa D, Logroscino G, Imbimbo BP (2018) BACE inhibitors in clinical development for the treatment of Alzheimer's disease. *Expert Rev Neurother* 18(11):847–857. <https://doi.org/10.1080/14737175.2018.1531706>
 15. Burki T (2018) Alzheimer's disease research: the future of BACE inhibitors. *Lancet* 391(10139):2486. [https://doi.org/10.1016/S0140-6736\(18\)31425-9](https://doi.org/10.1016/S0140-6736(18)31425-9)
 16. Lerchner A, Machauer R, Betschart C, Veenstra S, Rueeger H, McCarthy C, Tintelnot-Blomley M, Jatón AL, Rabe S, Desrayaud S, Enz A, Staufenbiel M, Paganetti P, Rondeau JM, Neumann U (2010) Macrocyclic BACE-1 inhibitors acutely reduce Abeta in brain after po application. *Bioorg Med Chem Lett* 20(2):603–607. <https://doi.org/10.1016/j.bmcl.2009.11.092>
 17. Turkenburg JP, Lamers MBAC, Brzozowski AM, Wright LM, Hubbard RE, Sturt SL, Williams DH (2002) Structure of a Cys25->Ser mutant of human cathepsin S. *Acta Crystallogr D Biol Crystallogr* 58(Pt 3):451–455. <https://doi.org/10.1107/s0907444901021825>
 18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242. <https://doi.org/10.1093/nar/28.1.235>
 19. Hong L, Turner RT 3rd, Koelsch G, Shin D, Ghosh AK, Tang J (2002) Crystal structure of memapsin 2 (beta-secretase) in complex with an inhibitor OM00-3. *Biochemistry* 41(36):963–967. <https://doi.org/10.1021/bi026232n>
 20. Hattori Y, Kobayashi K, Deguchi A, Nohara Y, Akiyama T, Teruya K, Sanjoh A, Nakagawa A, Yamashita E, Akaji K (2015) Evaluation of transition-state mimics in a superior BACE1 cleavage sequence as peptide-mimetic BACE1 inhibitors. *Bioorg Med Chem* 23(17):5626–5640. <https://doi.org/10.1016/j.bmc.2015.07.023>
 21. Neumann U, Ufer M, Jacobson LH, Rouzade-Dominguez ML, Huledal G, Kolly C, Lüönd RM, Machauer R, Veenstra SJ, Hurth K, Rueeger H, Tintelnot-Blomley M, Staufenbiel M, Shimshek DR, Perrot L, Friauff W, Dubost V, Schiller H, Vogt B, Beltz K, Avrameas A, Kretz S, Pezous N, Rondeau JM, Beckmann N, Hartmann A, Vormfelde S, David OJ, Galli B, Ramos R, Graf A, Lopez Lopez C (2018) The BACE-1 inhibitor CNP520 for prevention trials in Alzheimer's disease. *EMBO Mol Med* 10(11):e9316. <https://doi.org/10.15252/emmm.201809316>
 22. van Zundert GCP, Hudson BM, de Oliveira SHP, Keedy DA, Fonseca R, Heliou A, Suresh P, Borrelli K, Day T, Fraser JS, van den Bedem H (2018) qFit-ligand reveals widespread conformational heterogeneity of drug-like molecules in X-ray electron density maps. *J Med Chem* 61(24):183–198. <https://doi.org/10.1021/acs.jmedchem.8b01292>
 23. Shiba T, Kametaka S, Kawasaki M, Shibata M, Waguri S, Uchiyama Y, Wakatsuki S (2004) Insights into the phosphoregulation of beta-secretase sorting signal by the VHS domain of GGA1. *Traffic* 5(6):437–448. <https://doi.org/10.1111/j.1600-0854.2004.00188.x>
 24. Hong L, Tang J (2004) Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry* 43(16):4689–4695. <https://doi.org/10.1021/bi0498252>
 25. Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith KJ, Stanway S, Stemp G, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 1: identification of novel hydroxy ethylamines (HEAs). *Bioorg Med Chem Lett* 18(3):1011–1016. <https://doi.org/10.1016/j.bmcl.2007.12.017>
 26. Beswick P, Charrier N, Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Gleave R, Hawkins J, Hussain I, Johnson CN, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Skidmore J, Soleil V, Smith KJ, Stanway S, Stemp G, Stuart A, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 3: identification of hydroxy ethylamines (HEAs) with nanomolar potency in cells. *Bioorg Med Chem Lett* 18(3):1022–1026. <https://doi.org/10.1016/j.bmcl.2007.12.020>
 27. Chaput L, Selwa E, Elisée E, Iorga BI (2019) Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations. *J Comput Aided Mol Des* 33(1):93–103. <https://doi.org/10.1007/s10822-018-0161-7>

28. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612. <https://doi.org/10.1002/jcc.20084>
29. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35:D198–D201. <https://doi.org/10.1093/nar/gkl999>
30. Machauer R, Laumen K, Veenstra S, Rondeau JM, Tintelnot-Blomley M, Betschart C, Jaton AL, Desrayaud S, Staufenberg M, Rabe S, Paganetti P, Neumann U (2009) Macrocyclic peptidomimetic beta-secretase (BACE-1) inhibitors with activity in vivo. *Bioorg Med Chem Lett* 19(5):1366–1370. <https://doi.org/10.1016/j.bmcl.2009.01.055>
31. Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith KJ, Stanway S, Stemp G, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 2: identification of hydroxy ethylamines (HEAs) with reduced peptidic character. *Bioorg Med Chem Lett* 18(3):1017–1021. <https://doi.org/10.1016/j.bmcl.2007.12.019>
32. Charrier N, Clarke B, Cutler L, Demont E, Dingwall C, Dunsdon R, East P, Hawkins J, Howes C, Hussain I, Jeffrey P, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2008) Second generation of hydroxyethylamine BACE-1 inhibitors: optimizing potency and oral bioavailability. *J Med Chem* 51(11):3313–3317. <https://doi.org/10.1021/jm800138h>
33. Sandgren V, Agback T, Johansson PO, Lindberg J, Kvarnström I, Samuelsson B, Belda O, Dahlgren A (2012) Highly potent macrocyclic BACE-1 inhibitors incorporating a hydroxyethylamine core: design, synthesis and X-ray crystal structures of enzyme inhibitor complexes. *Bioorg Med Chem* 20(14):4377–4389. <https://doi.org/10.1016/j.bmc.2012.05.039>
34. Hanessian S, Shao Z, Betschart C, Rondeau JM, Neumann U, Tintelnot-Blomley M (2010) Structure-based design and synthesis of novel P2/P3 modified, non-peptidic beta-secretase (BACE-1) inhibitors. *Bioorg Med Chem Lett* 20(6):1924–1927. <https://doi.org/10.1016/j.bmcl.2010.01.139>
35. Sealy JM, Truong AP, Tso L, Probst GD, Aquino J, Hom RK, Jagodzinska BM, Dressen D, Wone DWG, Brogley L, John V, Tung JS, Pleiss MA, Tucker JA, Konradi AW, Dappen MS, Toth G, Pan H, Ruslim L, Miller J, Bova MP, Sinha S, Quinn KP, Sauer JM (2009) Design and synthesis of cell potent BACE-1 inhibitors: structure-activity relationship of P1' substituents. *Bioorg Med Chem Lett* 19(22):6386–6391. <https://doi.org/10.1016/j.bmcl.2009.09.061>
36. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. *Proteins* 52(4):609–623. <https://doi.org/10.1002/prot.10465>
37. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455–461. <https://doi.org/10.1002/jcc.21334>
38. Pennington LD, Whittington DA, Bartberger MD, Jordan SR, Monenschein H, Nguyen TT, Yang BH, Xue QM, Vounatsos F, Wahl RC, Chen K, Wood S, Citron M, Patel VF, Hitchcock SA, Zhong W (2013) Hydroxyethylamine-based inhibitors of BACE1: P1–P3 macrocyclization can improve potency, selectivity, and cell activity. *Bioorg Med Chem Lett* 23(15):4459–4464. <https://doi.org/10.1016/j.bmcl.2013.05.028>
39. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H (2015) Calculation of binding free energies. In: Kukol A (eds) *Molecular modeling of proteins. Methods in molecular biology* (Methods and protocols), vol 1215. Humana Press, New York, NY, pp 173–209. https://doi.org/10.1007/978-1-4939-1465-4_9
40. Wang L, Deng Y, Knight JL, Wu Y, Kim B, Sherman W, Shelley JC, Lin T, Abel R (2013) Modeling local structural rearrangements using FEP/REST: application to relative binding affinity predictions of CDK2 inhibitors. *J Chem Theory Comput* 9(2):1282–1293. <https://doi.org/10.1021/ct300911a>
41. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2015) Pmx: automated protein structure and topology generation for alchemical perturbations. *J Comput Chem* 36(5):348–354. <https://doi.org/10.1002/jcc.23804>
42. Crooks GE (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E* 60(3):2721–2726. <https://doi.org/10.1103/PhysRevE.60.2721>
43. Shirts MR, Bair E, Hooker G, Pande VS (2003) Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys Rev Lett* 91(14):140601. <https://doi.org/10.1103/PhysRevLett.91.140601>
44. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
45. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Bioinf* 65(3):712–725. <https://doi.org/10.1002/prot.21123>
46. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113(26):9004–9015. <https://doi.org/10.1021/jp901540t>
47. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct Bioinf* 78(8):1950–1958. <https://doi.org/10.1002/prot.22711>
48. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14:71–73. <https://doi.org/10.1038/nmeth.4067>
49. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general Amber force field. *J Comput Chem* 25(9):1157–1174. <https://doi.org/10.1002/jcc.20035>
50. Yesselman JD, Price DJ, Knight JL, Brooks CL III (2012) MATCH: an atom-typing toolset for molecular mechanics force fields. *J Comput Chem* 33(2):189–202. <https://doi.org/10.1002/jcc.21963>
51. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell D (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31(4):671–690. <https://doi.org/10.1002/jcc.21367>
52. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935. <https://doi.org/10.1063/1.445869>
53. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092. <https://doi.org/10.1063/1.464397>
54. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577–8593. <https://doi.org/10.1063/1.470117>
55. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINC: a linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H)

56. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(014):101. <https://doi.org/10.1063/1.2408420>
57. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52(12):7182–7190. <https://doi.org/10.1063/1.328693>
58. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2016) Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. *Angew Chem Int Ed* 55:7364–7368. <https://doi.org/10.1002/anie.201510054>
59. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB, Carlson HA, Burley SK, Walters WP, Amaro RE, Feher VA, Gilson MK (2016) D3R grand challenge 2015: evaluation of protein-ligand pose and affinity predictions. *J Comput Aided Mol Des* 30(9):651–668. <https://doi.org/10.1007/s10822-016-9946-8>
60. Schrödinger LLC (2015) The PyMOL molecular graphics system, version 1.8.0.3
61. Hagberg A, Swart P, Schult DA (2008) Exploring network structure, dynamics, and function using NetworkX. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States)
62. Hunter JD (2007) Matplotlib: a 2D graphics environment. *Comput Sci Eng* 9(3):90–95. <https://doi.org/10.1109/MCSE.2007.55>
63. Surpateanu G, Iorga BI (2012) Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors. *J Comput Aided Mol Des* 26(5):595–601. <https://doi.org/10.1007/s10822-011-9526-x>
64. Colas C, Iorga BI (2014) Virtual screening of the SAMPL4 blinded HIV integrase inhibitors dataset. *J Comput Aided Mol Des* 28(4):455–62. <https://doi.org/10.1007/s10822-014-9707-5>
65. Martiny VY, Martz F, Selwa E, Iorga BI (2016) Blind pose prediction, scoring, and affinity ranking of the CSAR 2014 dataset. *J Chem Inf Model* 56(6):996–1003. <https://doi.org/10.1021/acs.jcim.5b00337>
66. Selwa E, Martiny VY, Iorga BI (2016) Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets. *J Comput Aided Mol Des* 30(9):829–839. <https://doi.org/10.1007/s10822-016-9983-3>
67. Selwa E, Elisée E, Zavala A, Iorga BI (2018) Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations. *J Comput Aided Mol Des* 32(1):273–286. <https://doi.org/10.1007/s10822-017-0054-1>
68. Gapsys V, de Groot BL (2017) Alchemical free energy calculations for nucleotide mutations in protein-DNA complexes. *J Chem Theory Comput* 13(12):6275–6289. <https://doi.org/10.1021/acs.jctc.7b00849>
69. Aldeghi M, Gapsys V, de Groot BL (2018) Accurate estimation of ligand binding affinity changes upon protein mutation. *ACS Central Sci* 4(12):1708–1718. <https://doi.org/10.1021/acscentsci.8b00717>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.