RSC Advances

REVIEW

View Article Online
View Journal | View Issue

Check for updates

Cite this: *RSC Adv.*, 2021, **11**, 11026

Pepsin-like aspartic proteases (PAPs) as model systems for combining biomolecular simulation with biophysical experiments†

Soumendranath Bhakat

Pepsin-like aspartic proteases (PAPs) are a class of aspartic proteases which shares tremendous structural similarity with human pepsin. One of the key structural features of PAPs is the presence of a β-hairpin motif otherwise known as flap. The biological function of the PAPs is highly dependent on the conformational dynamics of the flap region. In apo PAPs, the conformational dynamics of the flap is dominated by the rotational degrees of freedom associated with $\chi 1$ and $\chi 2$ angles of conserved Tyr (or Phe in some cases). However it is plausible that dihedral order parameters associated with several other residues might play crucial roles in the conformational dynamics of apo PAPs. Due to their size, complexities associated with conformational dynamics and clinical significance (drug targets for malaria, Alzheimer's disease *etc.*), PAPs provide a challenging testing ground for computational and experimental methods focusing on understanding conformational dynamics and molecular recognition in biomolecules. The opening of the flap region is necessary to accommodate substrate/ligand in the active site of the PAPs. The BIG challenge is to gain atomistic details into how reversible ligand binding/unbinding (molecular recognition) affects the conformational dynamics. Recent reports of kinetics (K_i, K_d) and thermodynamic parameters (ΔH, $T\Delta S$, and ΔG) associated with macro-cyclic ligands bound to BACE1 (belongs to PAP family) provide a perfect challenge (how to deal with big ligands with multiple torsional angles and select optimum order parameters to study reversible ligand binding/unbinding) for computational methods to predict binding free energies and kinetics beyond typical test systems *e.g.* benzamide–trypsin. In this work, i reviewed several order parameters which were proposed to capture the conformational dynamics and molecular recognition in PAPs. I further highlighted how machine learning methods can be used as order parameters in the context of PAPs. I then proposed some open ideas and challenges in the context of molecular simulation and put forward my case on how biophysical experiments *e.g.* NMR, time-resolved FRET *etc.* can be used in conjunction with biomolecular simulation to gain complete atomistic insights into the conformational dynamics of PAPs.

Received 8th December 2020
Accepted 21st February 2021

DOI: 10.1039/d0ra10359d

rsc.li/rsc-advances

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Introduction

Aspartic proteases are a class of enzymes which consist of two highly conserved aspartates in the active site. These enzymes use an active water molecule bound (by H-bond interaction) to aspartate for catalysis of their substrates. Based on their three-dimensional structural similarity, aspartic proteases can be categorised into two categories: (1) pepsin-like aspartic proteases (PAPs) and (2) retroviral aspartic proteases (RAPs).[1] PAPs belong to the A01 family whereas RAPs belong to the A02 family of proteases within the MEROPS database[2] (Table 1).

In recent years, HIV protease (belongs to RAPs) gained significant attention in context of structural and computational biology due to its role in HIV/AIDS.[3] Similar argument can be put forward for β-secretase enzyme, BACE1 which belongs to PAPs and it is a promising drug target for Alzheimer's disease. One of the comprehensive reviews‡ on PAPs have been published by Ben Dunn in 2002.[1] Since then several blockbuster drug molecules have been proposed targeting BACE1.[4] However, conformational dynamics and molecular recognition in PAPs have not been extensively studied by biophysical experiments *e.g.* NMR, single molecule FRET, Raman spectroscopy *etc.* Recent computational studies on BACE1 and plasmepsins (malarial PAP) have identified key order parameters (dynamical

Division of Biophysical Chemistry, Center for Molecular Protein Science, Department of Chemistry, Lund University, P. O. Box 124, SE-22100 Lund, Sweden. E-mail: soumendranath.bhakat@bpc.lu.se; bhakatsoumendranath@gmail.com; Tel: +46-769608418

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0ra10359d

‡ Focusing on structural and mechanistic aspects of PAPs.

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 1 Structures (PDB IDs) and members of a few representative aspartic proteases. Abbreviations: BACE: β-secretase, HIV: human immunodeficiency viruses, SIV: simian immunodeficiency virus

PAPs (MEROPS: A01)	PDBs	RAPs (MEROPS: A02)	PDBs
Human pepsin	1PSN, 1QRP, 1PSO	HIV-1	4L1A, 1TW7, 4NKK, 2PC0
Human renin	1RNE, 1HRN, 3D91	HIV-2	1IDA
BACE-1	1W50, 3TPL, 1SGZ	SIV	1TCW, 1AZ5, 1SIP
BACE-2	3ZKG, 3ZKI, 2EWY		
Cathepsin D	1LYA, 1LYB		
Plasmepsin I	4CKU, 3QRV		
Plasmepsin II	1LF4, 1LF3, 4CKU, 4Z22, 1SME, 2BJU		
Plasmepsin IV	1LS5		
Plasmepsin V	4ZL4, 6C4G		
Plasmepsin IX	N/A		
Plasmepsin X	N/A		
Penicillopepsin	1APW, 1PPM, 1PPK, 1APU		

structural properties) that can be further explored by biophysical experiments combined with state-of-the art molecular simulation.[5–10] The purpose of this review is to highlight common order parameters that governs conformational dynamics and molecular recognition of PAPs and to highlight the need of integrated computational and experimental approaches to gain atomistic insights into molecular mechanism of PAPs. I will also highlight how machine learning methods can be used for this purpose and will put forward my case on using PAPs as a test system for methodological development in computational and experimental frontiers. In this review, i will use plasmepsins (Plms) and BACE1 as test systems to guide the discussion. PAPs share incredible structural similarity, hence i believe computational/experimental methodologies applied on the aforementioned proteins will be broadly applicable to other enzymes of this class. Biophysical experiments, machine learning and biomolecular simulation are the three pillars which governs drug discovery in 21st century.[11] I will further discuss how structural similarity of PAPs can be exploited in the context of drug re-purposing towards plasmepsins (drug target for malaria).

Structural features

The main aim of this section is to clarify the language, organize the concepts and define structural features across a few representative PAPs. Common sequential features that are conserved in PAPs are as follows: (1) presence of two catalytic aspartates (often known as catalytic site) and (2) the presence of –Asp–Thr–Gly–Ser– sequence (Fig. 1). Structurally PAPs can be characterised by two main domains (1) the flap region and (2) the coil region.[3,5,6] The name coil region is not accurate. The so called coiled region consists of β-strand and coil structures. In this review, it will be denoted as CS region (coil-strand region) (Fig. 2 and Table 2).

The flap region acts as a lid over the catalytic site which governs the entry of the substrate or small molecule inhibitors. The extent of flap opening (described in detail in later sections) also dictates the volume of the binding site *i.e.* open flap

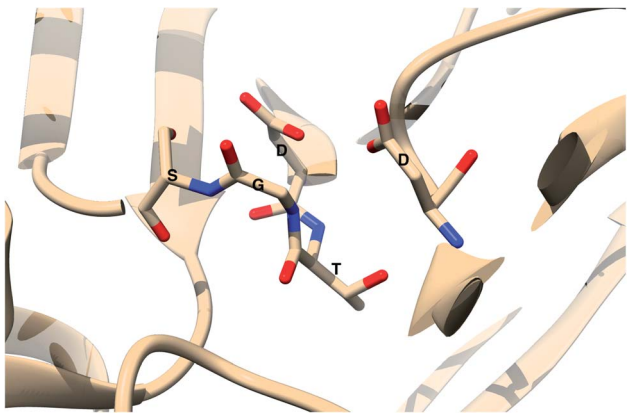

Fig. 1 Common structural features in PAPs: location of two conserved aspartates and the presence of –Asp–Thr–Gly–Ser– sequence. Figure was created using PDB:3TPL.[12] Amino acids are highlighted as one letter codes.

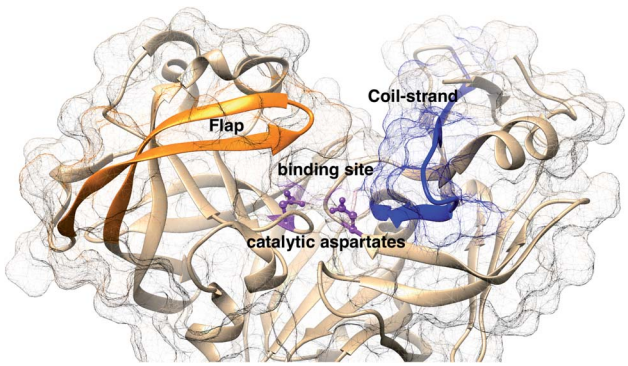

Fig. 2 Flap (resid 67–88) and coil-strand (resid 282–302) regions are two of the common structural features of PAPs. The flap acts as a lid over the catalytic aspartates. Opening of the flap exposes the binding site which enables ligand binding. Figure was created using PDB:1LF4.[13]

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 2 Residue numbers of the flap and CS region in PlmII and BACE1

Protein	Flap	CS region
PlmII	58–88	282–302
BACE1	52–82	314–334

conformation increases the volume of the binding site which can accommodate bigger ligands. The CS domain also plays a role in ligand/substrate binding but its role is not as pronounced as the flap region.

One of the startling structural features of PAPs is the presence of conserved tyrosine and tryptophan residue. The tyrosine (Tyr) residue is highly conserved in the flap region of most of the PAPs. Tryptophan (Trp) is another conserved residue that is present in the PAPs. In most cases, Trp is present at the S1 region (Fig. 3). However in case of BACE1 and BACE2, the conserved Trp is a part of the flap region (Fig. 3). Plasmepsin V (PlmV), an emerging drug target against malaria doesn't possess the conserved Trp residue.[14] Whereas in PlmIX and X, the conserved Tyr residue is replaced by phenylalanine (Phe) (Fig. 3). Visual inspection of several 3D crystal structure of Plm and BACE-1 shows that the orientation of Tyr and Trp influences the extent of flap opening (Fig. 4). In the next section, i will propose/review several order parameters that can be used to study the extent of flap opening and its connection with orientation of Tyr and Trp.

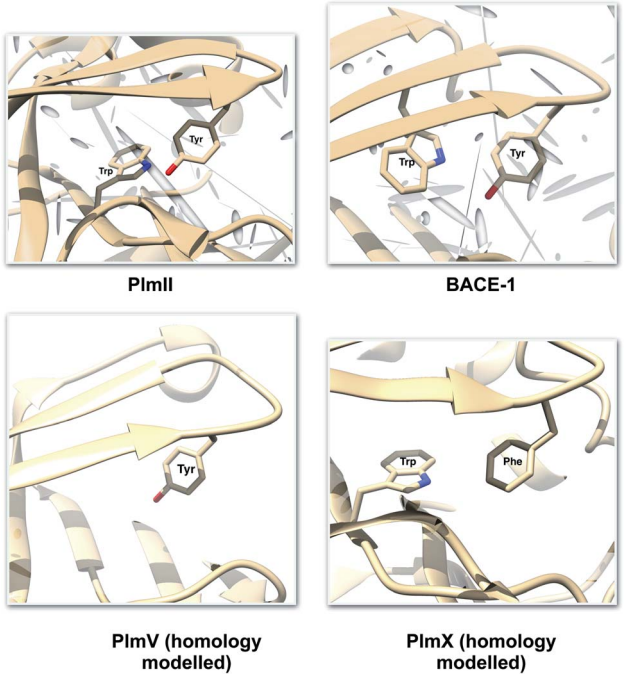

Fig. 3 Location of conserved Tyr and Trp. For most of the PAPs, the Trp belongs to the S1 region (substrate binding region 1) as depicted in PlmII. Whereas in case of BACE1, the Trp is a part of the flap region. Exceptions: PlmV doesn't possess the conserved Trp and in case of PlmIX/X the Tyr is replaced by Phe. The homology modelled structures were generated using SwissModel[15] and can be accessed in the corresponding Github profile.

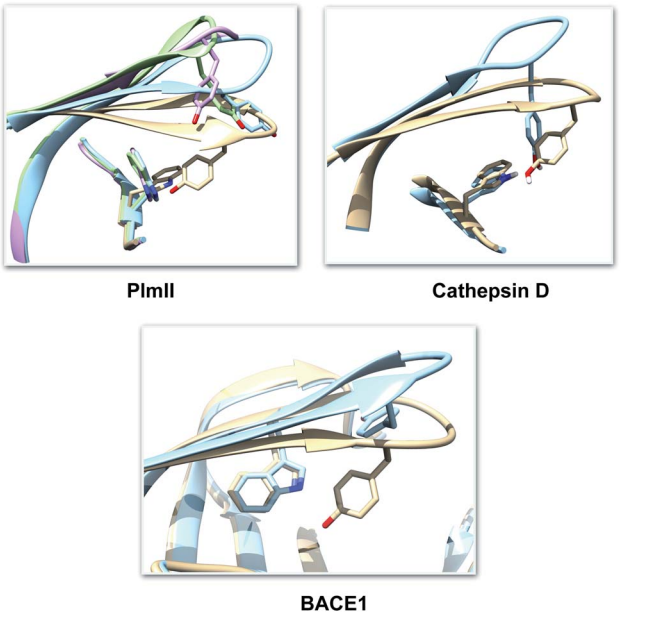

Fig. 4 Visual inspection of 3D structures of PAPs shows how orientation of Tyr and Trp alters flap elevation. In case of PlmII, altered Trp and Tyr orientation (PDB:2BJU,[16] 2IGY[17] and 4Z22 (ref. 18)) leads to expansion of the binding site (grey: 1LF4, blue: 2BJU, green: 2IGY, magenta: 4Z22). In case of cathepsin-D, high pH alters the conformation of Tyr which leads to flap elevation (blue: 1LYW,[19] grey: 1LYA[20]). Alternative conformation of Tyr was also observed in case of BACE-1 (blue: 1SGZ, grey: 3TPL).

Order parameters to study conformational dynamics

Order parameter is a commonly used term in physical chemistry. A poor man's working definition of order parameter (sometimes referred as reaction co-ordinate) in context of biomolecular conformational dynamics can be expressed as follows: 'geometric or abstract co-ordinate systems which captures conformational changes along a pathway'. In biomolecular simulation, order parameters are often known as collective variables (CVs). Some of the commonly used geometric order parameters are distance between two atoms, bond lengths, torsional angles *etc.* Abstract order parameters include vectorised linear combinations *e.g.* principal components. Time independent components, latent variable from variational auto-encoder *etc.* In this section, i will discuss how geometric and abstract order parameters can be used to capture conformational dynamics and molecular recognition in PAPs.

Geometric order parameters

Karubiu *et al.* proposed the Cα–Cα distance (DIST2) between catalytic Asp and the flap tip residue (flap tip residue varies across different PAPs) in PlmII as order parameters to capture the extent of flap opening in unbiased all-atom molecular dynamics (MD) simulation.[5]

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

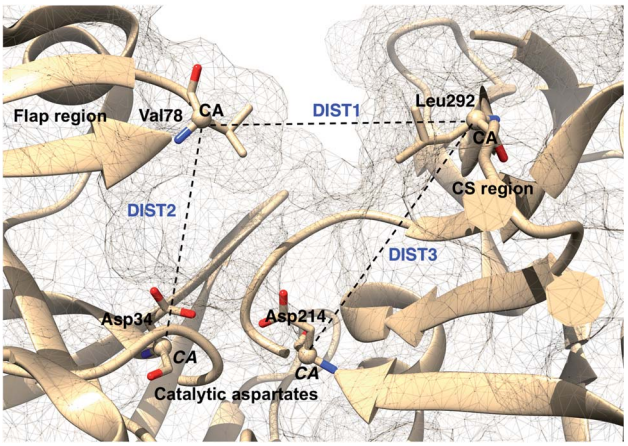

Fig. 5 Different distance based order parameters proposed by Karubiu *et al.* DIST2 measures the extent of flap opening whereas DIST3 accounts for opening of CS region. Fig. 12 shows that DIST2 and DIST3 are uncorrelated which gives an indication that in apo PAPs, the dynamics of flap and CS region are independent of each other. DIST1 acted as an auxiliary order parameter. These distance based order parameters can be generalised for other PAPs. The image was created using PDB:1LF4.

Rule of thumb to select flap tip residue. Residue number of the conserved Tyr/Phe + 1. For example, in PlmII it will be Val78 (77(Tyr) + 1).

This distance criteria was inspired by the order parameter proposed to study spontaneous flap opening and closing in apo HIV protease.[21,22] Karubiu *et al.* further proposed a distance metric (DIST3) which can capture opening/closing motion associated with CS domain (Fig. 5). Recent study by Bhakat and Söderhjelm[6] showed that the fluctuation of these two order parameters (DIST2 and DIST3) are uncorrelated which led to the hypothesis that in apo PAPs, the conformational dynamics of flap and CS domains are independent of each other (Fig. 12). Both DIST2 and DIST3 were further applied on BACE1 and other plasmepsins in order to gain insights into conformational dynamics from classical MD simulations. Bhakat and Söderhjelm performed test calculations (not published) where they have used DIST2 and DIST3 as CVs within metadynamics framework. The authors observed that the metadynamics biasing along these distances led to distorted flap conformations which suggests that these distance based order parameters are not the optimum choices to capture reversible conformational dynamics in PAPs but rather can be used as a quantitative measurement to capture the extent of flap/CS opening/closing. The authors further proposed two additional distance based order parameters COM_{flap} and COM_{CS} in case of PlmII with an aim that both these order parameters will simultaneously act as quantitative measurement of the conformational dynamics and CVs in enhanced sampling calculations *e.g.* metadynamics.

COM_{flap}: distance between centre of mass (COM) of the catalytic aspartates (only Cα atoms of catalytic aspartates) and COM of the flap region (Cα atoms of residues 58–88).

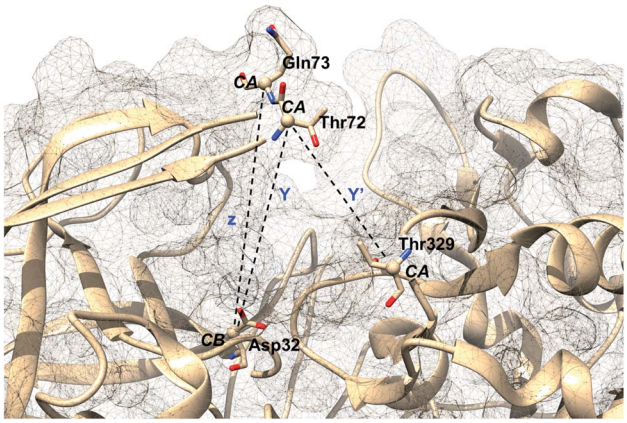

Fig. 6 Distance based order parameters proposed by other research groups in context of BACE1. One can easily see that *Y* and *z* are analogous to DIST2 whereas *Y'* is analogous to DIST1 (Fig. 5). The image was created using PDB:1W50.[24]

COM_{CS}: distance between COM of the catalytic aspartates (only Cα atoms of catalytic aspartates) and COM of the CS region (Cα atoms of residues 282–302).

Authors further used COM_{flap} and COM_{CS} as CVs in two – dimensional well-tempered metadynamics simulation. Metadynamics simulations with COM CVs showed significant hysteresis and lack of reversible sampling of the conformational landscape as these CVs doesn't represent slow order parameters (more on this later) necessary for optimum conformational sampling (Fig. 13).

Karubiu *et al.* further proposed a distance based order parameter (Fig. 5) which measures the distance (DIST1) between flap and CS region as a function of time. Due to uncorrelated dynamics of these regions, DIST1 doesn't provide any additional information compared to DIST2/DIST3. A few other distance based order parameter has also been proposed by different research groups. Gorfe and Caflisch[9] proposed (in

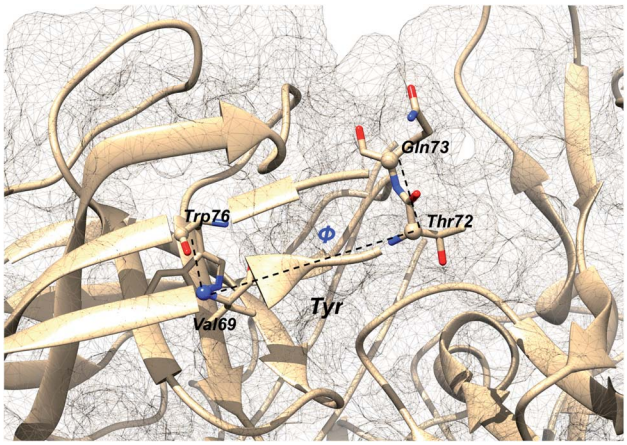

Fig. 7 Dihedral order parameter, ϕ proposed by Spronk and Carlson to account for flap twisting. The position of conserved Tyr is highlighted in order to give an idea on the relative position of other residues. The image was created using PDB:1W50.

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

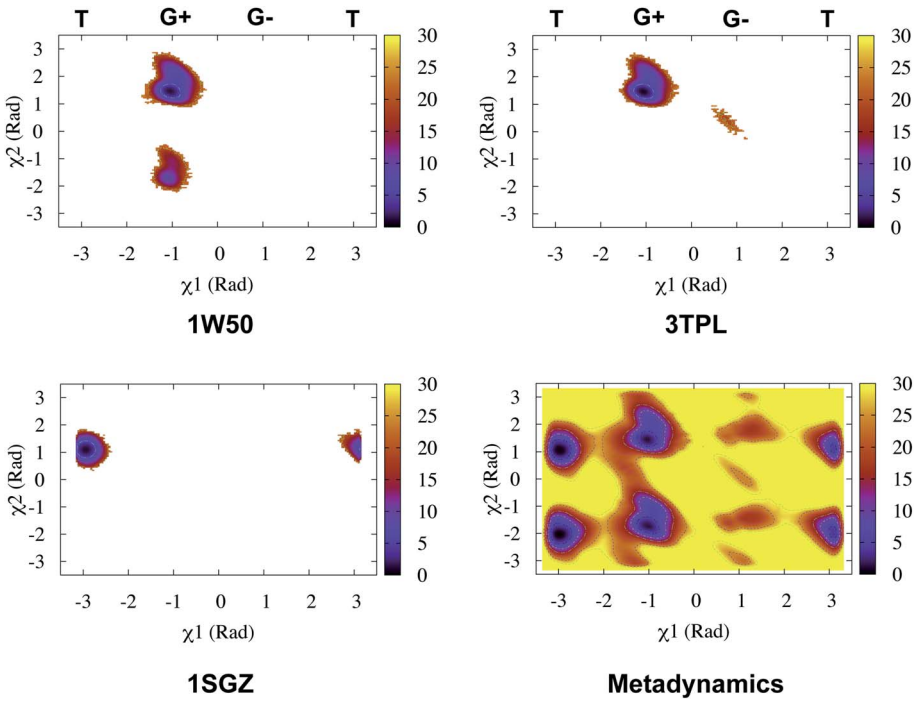

Fig. 8 Sampling of $\chi1$ and $\chi2$ angles in classical MD simulations (~ 500ns) starting with three different starting conformations of BACE1 (PDB ID: 1W50, 3TPL and 1SGZ differs in flap elevation and orientation of Tyr. For detail refer[6]). As one can see simulations starting with 1W50 and 3TPL only sampled *gauche*$^+$ (G^+) conformation. Whereas, simulation starting with 1SGZ only sampled *trans* (T) conformation. Metadynamics simulation with $\chi1$ and $\chi2$ order parameters did a enhanced sampling of the conformational space. Looking at apparent free energy values (z axis unit in kJ mol^{-1}) tells us that in apo BACE1 the flap remains in dynamic equilibrium between *gauche*$^+$ and *trans* conformations. *Gauche*$^-$ (G^-) is the minor population which was stabilised by H-bond interaction between Tyr and Lys.

context of BACE1) the distance between the Cα of the flap tip residue Thr72 and Cβ of the catalytic aspartate as a measure of flap opening ($\boldsymbol{Y}$). This order parameter is almost identical to the DIST2 proposed by Karubiu *et al.* The authors further proposed an analogous distance criteria of DIST1: distance between Cα atoms of flap tip Thr72 and Thr329 of CS region ($\boldsymbol{Y'}$). Spronk and Carlson[10] introduced an analogous distance parameter (in BACE1) of DIST2 (denoted as $\boldsymbol{z}$) which measures the distance between the Cα of the Gln73 and the Cβ of Asp32 (Fig. 6). Due to their similarity, either DIST2 or z can be applicable in PAPs to capture the extent of flap opening. Recently Shen and co-workers[23] proposed the following distance parameters to measure the extent of flap opening in human renin: (1) distance between Tyr-83-OH and Asp-38-CG and (2) distance between Ser-84-CB and Asp-226-CG (Fig. 27 in ESI†). The distance between Tyr-83-OH and Asp-38-CG can be a deceptive measurement of the flap opening as the orientation of Tyr-83-OH is highly dependent on the rotational degrees ($\chi1$ and $\chi2$) of freedom of Tyr. Flipping (discussed later) of Tyr side-chain can change the orientation of the Tyr-83-OH. In that case, the distance between Tyr-83-OH and Asp-38-CG will not measure the true extent of flap opening. Distance between Ser-84-CB and Asp-38-CG consist of two stable anchor points which can act as an alternative to Tyr-83-OH and Asp-38-CG.

Additional order parameters have been proposed to capture flap twisting. Spronk and Carlson proposed flap twisting angle, ϕ as the dihedral angle involving Trp76C–Val69N–Thr72CA–Gln73CA (Fig. 7). However, this definition is specialised for BACE1. A general order parameter for capturing flap twisting in PAPs can be expressed as a dihedral angle involving following residues:

$i + 5 - \text{C}–i - 2 - \text{N}–i + 1 - \text{CA}–i + 2 - \text{CA}$ where i is the residue number of conserved Tyr.

Recently, Bhakat and Söderhjelm proposed the torsion angles ($\chi1$ and $\chi2$) of conserved Tyr (in PlmII and BACE1) as order parameters to capture conformational dynamics in PAPs.[6] A typical free energy profile of Tyr (along $\chi1$) has three free energy minimas which corresponds to three different side-chain orientations: *gauche*$^+$, *gauche*$^-$ and *trans*. MD and metadynamics investigations using PlmII and BACE1 predicted that in apo PAPs both *gauche*$^+$ and *trans* conformations are equally populated§ (Fig. 8 and 9). *Gauche*$^+$ is stabilised by side-chain H-bond interaction between Tyr and Trp (denoted as normal state) whereas the *trans* conformation is stabilised by side-chain H-bond interaction between Tyr and one of catalytic Asp (denoted as flipped state, recently Shen and co-workers sampled this interaction using constant pH molecular dynamics on human renin) (Fig. 10). *Gauche*$^-$ predicted to be the minor population which in case of BACE1 is stabilised by side-chain H-bond

§ **Caution:** due to lack of experimental validation on the varied population of different Tyr conformations in apo PlmII/BACE1, the readers should take the computational prediction with skepticism. The computational predictions are susceptible to choice of water-models, force-fields and protonation states.

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

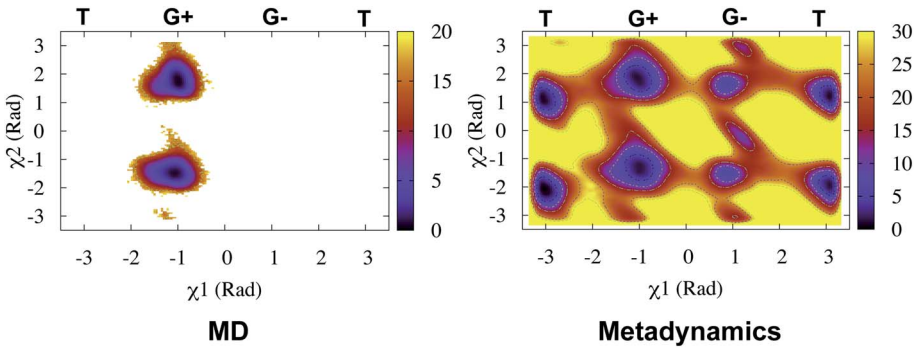

Fig. 9 Sampling of $\chi 1$ and $\chi 2$ order parameters in MD and metadynamics simulations starting with apo PlmII (PDB ID:1LF4). Sampling of MD and metadynamics was quite similar to Fig. 8.

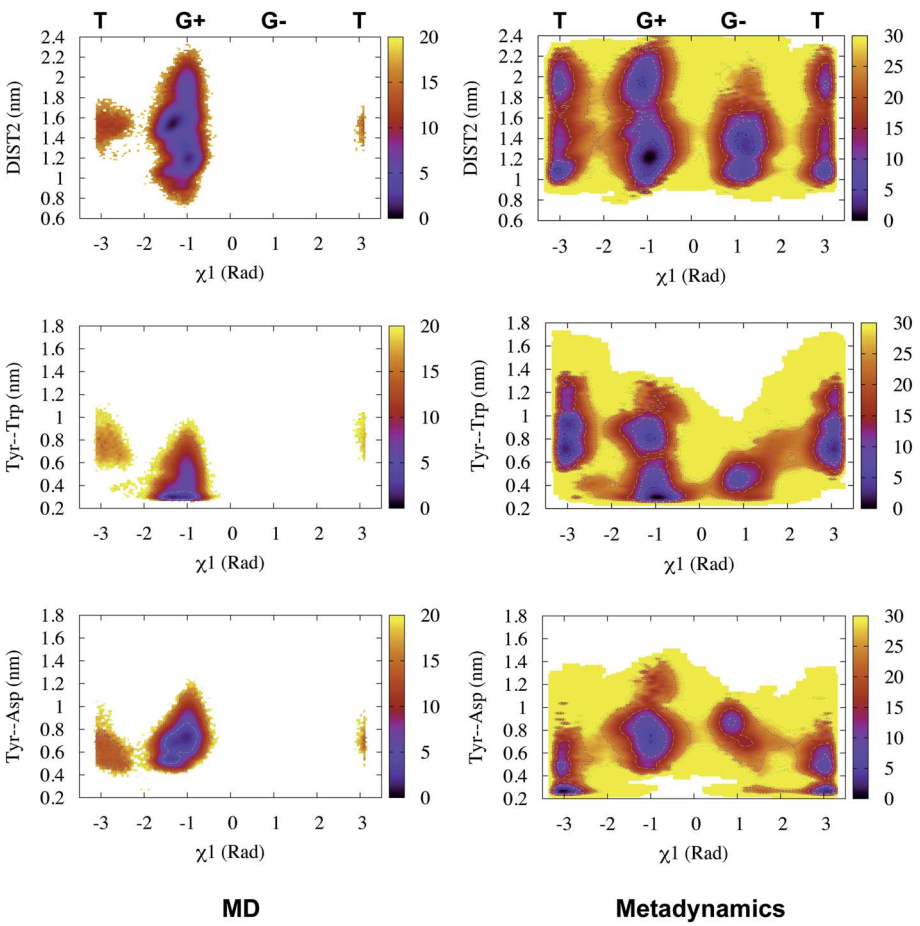

Fig. 10 Sampling of $\chi 1$ and DIST2 order parameters in MD (~600ns) and metadynamics simulations starting with apo PlmII (PDB ID:1LF4). The open flap conformation centres around DIST2 ~ 2.0 nm. The *gauche*$^+$ conformation was stabilised by H-bond interaction between Tyr and Trp whereas the *trans* conformation was stabilised by H-bond interaction between Tyr and catalytic Asp. The typical H-bond distance is 0.3–0.35 nm. *z* axis in the contour plot shows apparent free energy in kJ mol^{-1}.

interaction between Tyr and Lys (PDB:1SGZ[25]) (Fig. 4 and 8). A typical way to understand the rotation of Tyr and its connection with flap elevation is to plot a free energy profile involving DIST2/*z* against $\chi 1$ (Fig. 10).¶

Investigation of crystal structures of plasmepsin have shown that not only Tyr but the conserved Trp can also adapt flipped orientation (Fig. 11). Flipping of Trp (Table 3) combined with flap opening (measured using DIST2, see Table 4) leads to expansion of the binding pocket which can accommodate open flap inhibitors. The interplay between flap opening and dihedral order parameters of Tyr and Trp have further being

¶ One can make a cos transformation of $\chi 1$ which will handle the periodicity associated with χ angles.

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 3 Dihedral order parameters $\chi 1$ and $\chi 2$ (values in radian) shows different conformational states of Trp in Plm-II

PDBs	$\chi 1$	$\chi 2$
1LF4	1.22	−1.74
2BJU	−1.04	1.43

Table 4 Crystal structures of Plasmepsin-II (PDB:1LF4, 2BJU, 2IGY, 4Z22), bovine chymosin protease (PDB:1CMS[28]), human cathepsin-D (1LYA, 1LYW) and BACE-1 (3TPL, 1W50, 1SGZ) show different conformational states of Tyr. Crystallographic structures of these PAPs shows the variation in flap opening (DIST2). In case of human cathepsin-D, pH alters the extent of flap opening and the orientation of Tyr

PDB	χ_1 (radian)	DIST2 (nm)	State
1LF4	−1.12	1.19	*Gauche*$^+$
2BJU	−3.12	1.50	*Trans*
2IGY	−2.6	1.74	*Trans*
4Z22	−1.02	2.11	*Gauche*$^+$
1CMS	3.08	1.28	*Trans*
1LYA	−1.29	1.20	*Gauche*$^+$
1LYW	0.96	1.78	*Gauche*$^-$
3TPL	−1.31	1.23	*Gauche*$^+$
1W50	−1.06	1.76	*Gauche+*
1SGZ	0.87	1.61	*Gauche−*

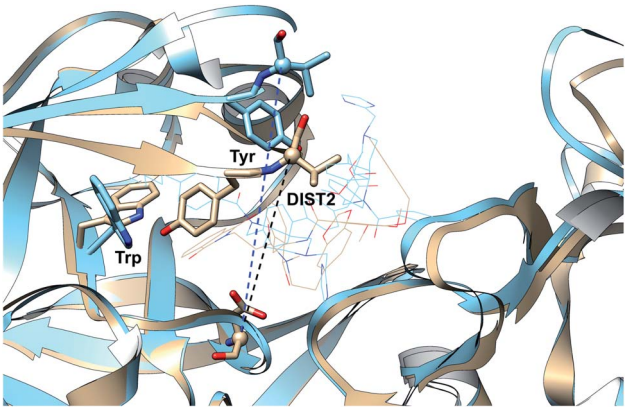

Fig. 11 Flipped orientation of Tyr and Trp in PlmII which led to flap opening (measured by DIST2) and expansion of the binding site. The expanded binding site can accommodate bigger ligands (represented as wire-frames). These inhibitors are often dubbed as open flap inhibitors.[27] In this case the rotation of Trp side-chain opens up space so that it can accommodate the alkane side-chain of non-peptido-mimetic inhibitor. Blue represents PDB ID:2BJU and grey represents PDB ID:1LF3.[13]

exploited by Oefner and colleagues in human renin.[26] By increasing flap elevation by 1 Å, rotating Tyr by −120° (flipped conformation) and displacing Trp slightly led to expansion of the binding site‖ which led to identification of novel piperidine based inhibitors with micro-molar k_i value. Further computational and experimental studies (X-ray crystallography and NMR) are necessary to understand if the combined flipping of Tyr and Trp is rare or a common phenomena in PAPs that correlates with flap opening and expansion of the binding site.

‖ Lack of structural reports (PDB structures) on this particular system makes it difficult to cross validate the claims.

Mutation of Tyr and its effect on conformational dynamics

Mutation of Tyr to Ala (amino acid with no bulky side-chain) in apo PlmII and BACE1 led to complete flap collapse in MD simulation which was captured by DIST2 (Fig. 12). This observation is consistent with previous experimental study by Suzuki and co-workers[29] which shows that Tyr to Ala mutation resulted in loss of enzyme activity in human renin. Further, mutation of Tyr to Thr, Ile, Val in chymosin also led to loss of enzyme activity. MD simulation based mutational study together with experiments led us to conclude that the conserved Tyr plays a critical role in enzyme activity. In future, similar experiments on PlmII and BACE1 are necessary to generalise the aforementioned claim.

Substitution of Tyr to Phe in pepsin retains the enzyme activity. Structural investigation of *Toxoplasma gondii* PAP, PlmIX, PlmX and PlmV reveals the presence of Phe in place of Tyr (Fig. 4). Since Phe possess rotational degrees of freedom along $\chi 1$ and $\chi 2$ order parameters hence it dictates the flap dynamics in the aforementioned PAPs. In future, this assumption can be validated by mutating Phe to Ala which results in loss of enzyme activity and flap collapse in MD simulation.

Mutation of Trp

Mutation of Trp and its effect on enzyme activity of PAPs has not been studied extensively. Park *et al.*[30] mutated conserved Trp (Trp39) of *R. pusillus* PAP with other residues and observed decreased enzyme activity. However, crystal structures of PlmV[14] from *P. vivax* and *P. falciparum* (homology modelled) doesn't possess conserved Trp. In future detailed computational and experimental studies are required in order to decipher the exact role of Trp in conformational dynamics of PAPs.

Abstract order parameters

MD simulations of PAPs can produce high-dimensional dataset with million or more data points. These data points in together describes conformational motion associated with protein. Recently several dimensionality reduction techniques (*e.g.* PCA,[31–33] tICA,[34–38] tSNE,[39–42] diffusion map[43,44]) have been proposed which can extract useful informations related to conformational dynamics from the plethora of the data generated during MD simulations.

Principal component analysis

Principal component analysis (PCA) does a linear transformation which (in context of MD simulation) aims at finding motions that maximize the variance. Before I dive into how it can be used as an order parameter to capture conformational dynamics associated with PAPs, I will briefly introduce the fundamental concept behind PCA.

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

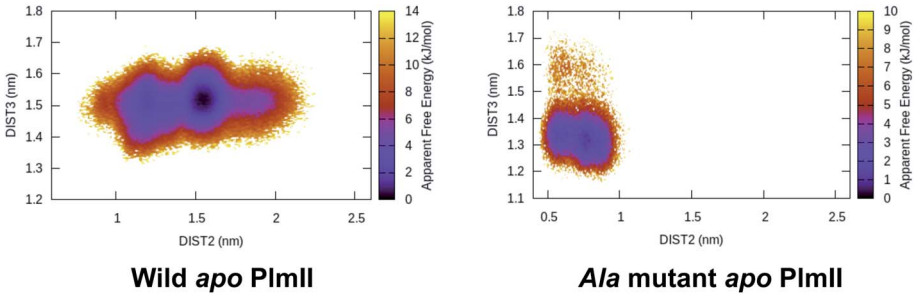

Fig. 12 Tyr to Ala mutation shows complete flap collapse in apo PlmII as captured by DIST2 in MD simulation. Similar results were obtained in case of BACE1 (ref. 6). This observation is in agreement with experiments where Tyr to Ala mutation resulted in loss of enzyme activity.

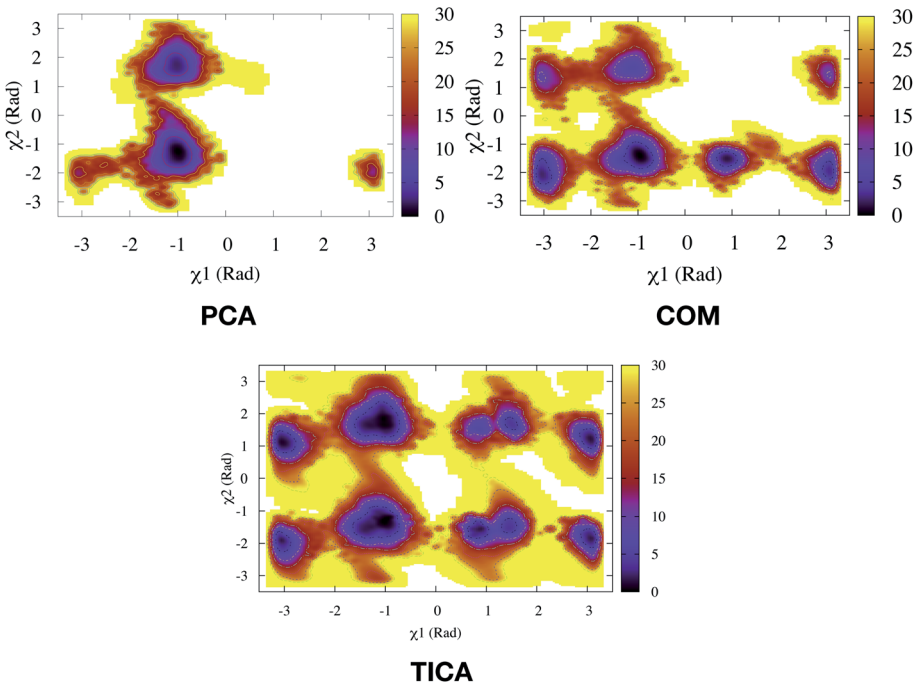

Fig. 13 Sampling of conformational space of apo PlmII using PC1 and PC2 as CVs in metadynamics (PCA). The authors also performed metadynamics simulations with COM CVs (COM). The sampling can be compared with metadynamics with χ1 and χ2 CVs (Fig. 9).Both PCA and COM based metadynamics didn't explicitly bias the torsional order parameters of conserved Tyr. Well-tempered metadynamics simulation with TIC1 and TIC3 as order parameters (TICA) did better sampling compared to PCA and COM CVs (ref. 6).

The first principal component (often denoted as PC1, PCA1 or z_1) is the linear trans formation of the original variables x_1, x_2, …., x_p with α. Mathematically this transformation can be expressed as:

$$z_1 = \alpha_1^{\mathrm{T}} x = \alpha_{11} x_1 + \alpha_{12} x_2 + \ldots\ldots + \alpha_{1p} x_p = \sum_{j=1}^{p} \alpha_{1j} x_j \quad (1)$$

The weight components (α_{11}, α_{12}, …, α_{1p}) maximizes the variance of the original data **x** subject to the following constraint (normalization):

$$\alpha_{11}{}^2 + \alpha_{12}{}^2 + \ldots\ldots + \alpha_{1p}{}^2 = 1 \quad (2)$$

Generally speaking k^{th} PC can be expressed as $z_{\mathrm{k}} = \alpha_{\mathrm{k}}^{\mathrm{T}} x$,** has the maximum variance and uncorrelated with z_1, z_2, …, z_{k-1}. **x** is the eigenvector matrix. Cartesian coordinates for each time-step in a MD simulation defines each of the rows in **x** whereas, p columns of **x** consist of $3N$ cartesian co-ordinates for each atom. For detailed description of PCA in context of MD simulation please refer ref. 32 and[33].

PCA algorithm as a post processing tool for MD simulation have been implemented in several state-of-the art software packages *e.g.* Gromacs,[45] MSMBuilder,[46] MDTraj[47] *etc.* In practice, first few principal components capture the motions of maximal variance from MD trajectories. Bhakat and Söderhjelm performed two independent MD simulations (~500 ns) of PlmII

** The second principal component can be expressed as $z_2 = \alpha_2^{\mathrm{T}} x$

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

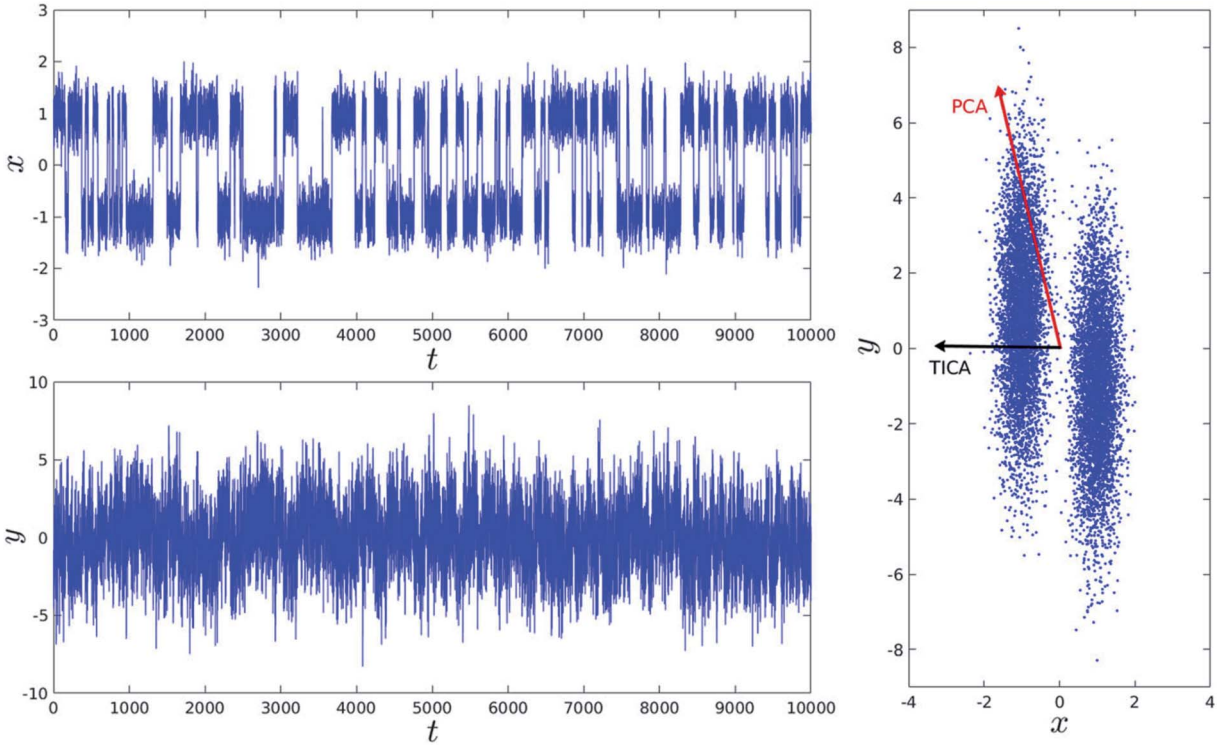

Fig. 14 Top and bottom plot shows time (t) series projection of two arbitrary order parameters, x and y. Right panel shows the dominant PCA and TICA vectors projected on x and y. Figure is adapted from ref. 50.

and performed PCA using the Cα atoms of the combined trajectory (ignoring the tail part) using g_{covar} tool integrated with Gromacs. The authors thought that first few PCs will capture degrees of freedom associated with the flap opening of PlmII. Metadynamics simulations using PC1 and PC2 as order parameters did a poor sampling of the conformational space of apo PlmII. PCA using the Cα atomic co-ordinates didn't take into account the dihedral angles associated with Tyr (as described in previous section) hence it is not a surprise that the metadynamics with PC1 and PC2 as order parameters failed to enhanced the sampling of the conformational space in apo PlmII (Fig. 13). An alternative approach will be to perform PCA on the dihedral (often known as *dihedral* PCA[31]) space (using χ1 and χ2 angles of the flap region in apo PlmII) which in principle should able to capture Tyr mediated flap dynamics in PAPs.

Independent component analysis

Second order independent component analysis (ICA) or otherwise known as time-lagged ICA (TICA) has also been applied in context of PAPs. While in PCA the aim was to find orthogonal linear transformations (PCs) that maximizes the variance, the goal in TICA is to identify linear transformations where the vectors (ICs) are statistically independent and maximizes autocorrelation (Fig. 14). ICA based approaches have been used widely to analyze time-series data from fMRI[48] and EEG[49] experiments. The first TIC component/vector (often denoted as TIC1) captures the slow (in context of timescale of biomolecular motions) dynamical mode from input high-dimensional time-series data (dihedral order parameters of the flap region in PAPs). TIC vectors can also be used as order parameters to perform well-tempered metadynamics simulations. Bhakat and Söderhjelm performed TICA analysis (for full mathematical description of TICA please refer[38] and[36]) on the MD trajectory of apo PlmII to identify key dihedral order parameters (more precisely linear combination of dihedral angles with different weights for each angle) that captures slow dynamical modes from MD trajectory (Fig. 15 and 16). The authors further used TIC vectors as order parameters in metadynamics simulation which enhances the sampling of flap dynamics compared to classical MD simulations (Fig. 13).

In principle analogous methods of TICA[51] *i.e.* VAC[52] and SGOOP[53] †† can also be applied on PAPs to identify and bias order parameters which capture slow dynamical modes (from the MD trajectory) corresponds to flap dynamics.

Integration of several ICA algorithms *e.g.* Infomax,[55] kernel ICA,[37] JADE,[57] Fast ICA[58] with programming interfaces such as Python opens up the door to test the effectiveness of these methods in identifying order parameters from the high-dimensional dihedral space of PAPs.

†† TICA and VAC are methodologically equivalent. Tiwary and coworkers have demonstrated that SGOOP follows principle of maximum entropy (MaxEnt).[53,54] An information theoretic approach for ICA has also highlighted its relation with MaxEnt[38,55,56]

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

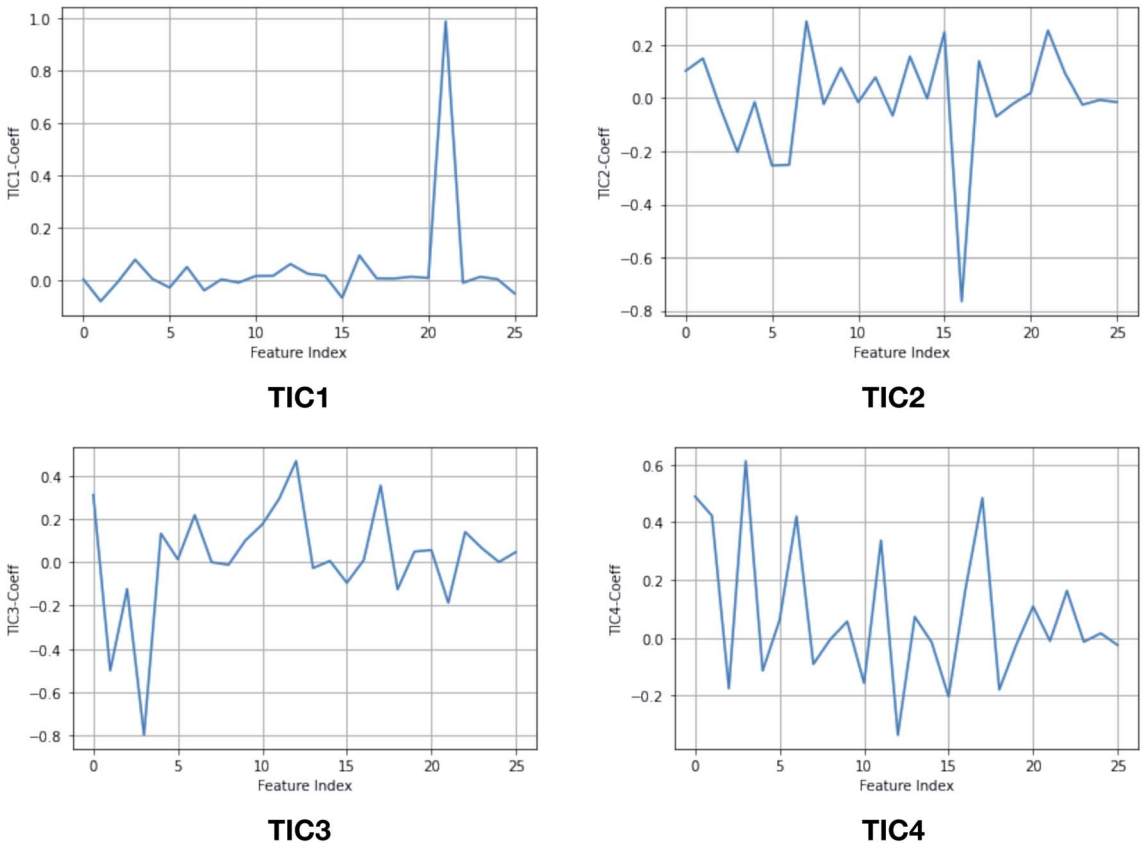

Fig. 15 TIC vectors and corresponding weights assign to each features. In case of TIC1, feature index 21 has significant weight which corresponds to sin $\chi 2$. Whereas in case of TIC3, feature index 3 has most significant weight which corresponds to sin $\chi 1$. 2D well-tempered metadynamics simulations with TIC1 and TIC3 as CVs did a better sampling of the conformational space of apo PlmII compared to classical MD simulation (Fig. 13). Note: technically it is possible to perform metadynamics simulations with all five TICs using parallel-bias metadynamics.

Binary classifiers

Binary classifiers are a set of supervised machine learning algorithms which aims at classifying an object into one of the two possible categories *e.g.* classifying the pictures of cats from dogs. The concept of using binary classifiers in automated selection of order parameters was first introduced by Sultan and Pande.[59] However, one prerequisite of using binary classifiers is that one has to sample the start and end states. In case of PAPs, these two states (otherwise known as state A and state B) can be normal and flipped states. If one extracts co-ordinates correspond to normal and flipped states from MD trajectory then the next step is to generate subset of order parameters (*e.g.* $\chi 1$ and $\chi 2$ angles of the flap region) which represents these two states. This is followed by training of the binary classifier *e.g.* support vector machine (SVM),[60] passive-aggressive (PasAg) classifier,[61] logistic regression,[62] perceptron[63] on the subset of order parameters. In case of PasAg classifier, the classifier decision boundary can be used as a order parameter to drive enhanced sampling calculation (Fig. 17). One advantage of using binary classifier based order parameters in metadynamics is that one can simultaneously bias multiple degrees of freedom (using parallel-bias metadynamics) which can lead to faster convergence (less chance for metadynamics to remain stuck along arbitrary orthogonal slow degrees of freedom) of metadynamics simulations. For example, if transition between normal to flipped state requires change in both $\chi 1$ and $\chi 2$ angles of Tyr and additional $\chi 1$ and $\chi 2$ angles of other amino acids (of the flap region), then binary classifier based order parameter would allow addition of metadynamics bias along these additional features using a single order parameter (classifier decision boundary). Initial calculations (unpublished) by the author have shown the effectiveness order parameters in sampling normal to flipped transition in apo PlmII (Fig. 18 and 19).

Another method that can be applied in capturing normal to flipped transition in PAPs is Fisher's linear discriminant analysis (LDA)[64] (Fig. 18). LDA has been routinely used in machine learning as a robust supervised classification method. Recently, Parrinello and co-workers have proposed a modified version of LDA, harmonic linear discriminant analysis (HLDA)[65] based order parameter which can be applied in context of normal to flipped transition in PAPs. However from a sampling point of view, i believe HLDA will not provide any additional benefits compared to binary classifier *e.g.* SVM, PasAg, logistic regression *etc.*

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

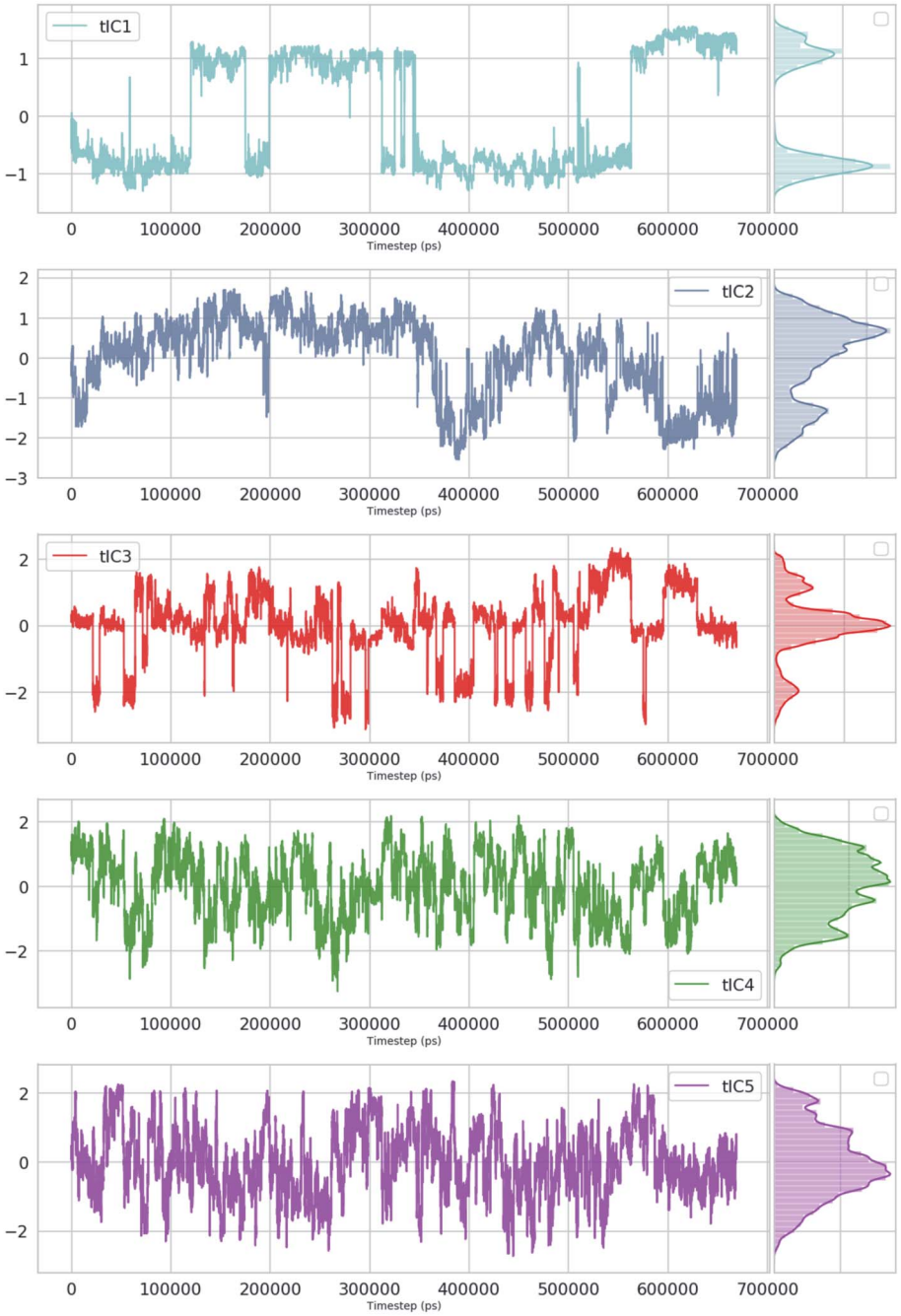

Fig. 16 TIC1–5 projected as a function of time for an unbiased ~600 ns MD simulation with TIP4P water model of apo PlmII.

PAPs: testing ground for computational method development

'If you know the answer beforehand it is easier to test new methods'

This was the philosophy behind using alanine dipeptide, chignolin, BPTI (together these systems known as the mice model for simulation methods) as typical test systems for novel computational methods directed towards order parameter selection. As discussed before the $\chi 1$ and $\chi 2$ order parameters[6] of Tyr plays a critical role in conformational dynamics of apo PAPs. The rotation degrees of freedom associated with Tyr is believed to be one of the slow degrees of freedom which governs conformational dynamics of the flap. However, rotation degrees of freedom associated with conserved Trp and other residues present in the flap region also plays crucial role in overall conformational dynamics of PAPs. High dimensionality associated with dihedral angles of PAPs makes it an ideal play ground to test neural network based latent variable order parameters. In mathematical terms, latent variables are transformations that are not directly observed (probability distributions of dihedral angles) but are rather emerged (using

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

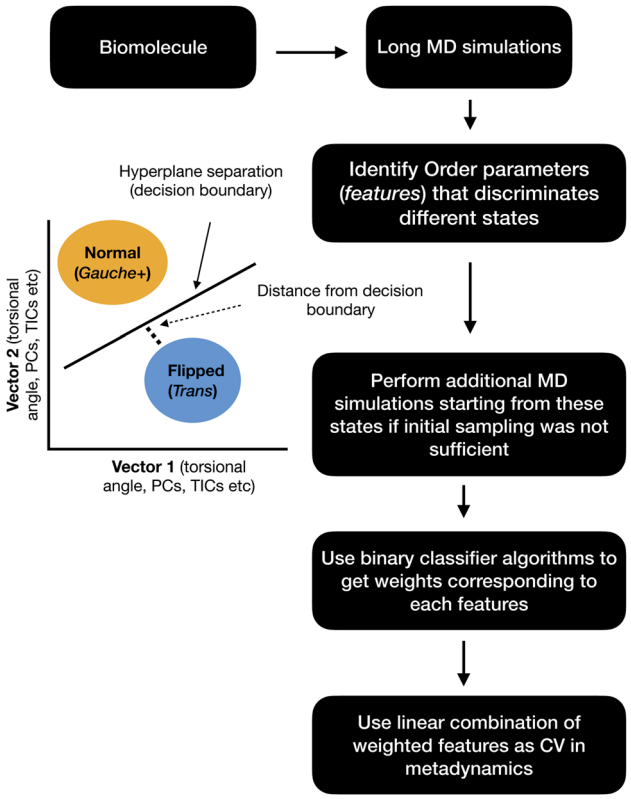

Fig. 17 Schematic diagram showing the steps involved in the application of binary classifier algorithms in context of biomolecular simulation. In case of PAPs vector1 and vector2 can be the $\chi1$ and $\chi2$ angles of the flap region. The $\chi1$ and $\chi2$ angles of Tyr plays dominating role in discriminating normal and flipped states. One can also use states separated along abstract order parameters (PCs, TICs) as classification problem.

mathematical transformations) from observed variables. Latent variable from bottleneck layer of neural network can effectively capture slow degrees of freedom from unbiased MD trajectories (Fig. 21 and 20). This type of neural network based order parameter can be further used to drive enhanced sampling calculations. Sidky *et al.*[66] and Wang *et al.*[67] laid out several machine learning algorithms that have been tested in context of biomolecular simulation. I believe most of these algorithms can be applied on PAPs with an aim to automatise identification of order parameters that governs its conformational dynamics. Bigger size of PAPs and prior knowledge on the role of conserved Tyr in flap dynamics (makes it easier to interpreted abstract order parameters) makes PAPs an ideal testing ground for novel machine learning algorithms (*e.g.* slow feature analyses,[68] LSTM[69] or transformer[70] like recurrent neural network).

Open idea

Can one design a deep learning algorithm which will take multiple X-ray structures as inputs and predict trial CVs that captures slow degrees of freedom which governs conformational dynamics of a biomolecule? One can use PAPs as a test case in this project. The variations in PDB structures of BACE1 and PlmII are mainly dominated by torsional order parameters associated with conserved Tyr. An output trial CV from a deep learning architecture should give higher weights to torsional order parameters associated with Tyr. Besides, it must capture some non-linear combinations of other torsional order parameters that vary among these X-ray structures. The trial CVs can be iteratively optimized using a metadynamics framework similar to VAC.[52]

Order parameters to capture ligand binding/unbinding

Ligand or substrate binding with PAPs require conformational change *i.e.* opening of the flap so that ligand/substrate can bind to the active site. Flap opening in PAPs is similar to the flap opening in HIV/SIV protease.[75] A generalised order parameter that can describe ligand binding/unbinding can be defined as follows:

'COM_{unbind}: the distance between COM of the ligand heavy atoms and the COM of the catalytic aspartates (Cα atoms)'

Plotting the aforementioned order parameter COM_{unbind} in combination of DIST2 gives an indication of the extent of flap opening during ligand binding/unbinding. Further, plotting COM_{unbind} with $\chi1$ and $\chi2$ angles of Tyr can give an idea on how flipping of conserved Tyr residue affects ligand binding/unbinding.

However, reversible sampling of ligand binding/unbinding using physical order parameters such as COM_{unbind} is not a trivial task. In one such initial effort Bhakat and Söderhjelm used COM_{unbind} order parameter within well-tempered metadynamics simulation. The results showed that flap opening is necessary for ligand unbinding (Fig. 22). However, the challenge is to multiple reversible sampling of ligand binding/unbinding in order to validate the role of flap opening with statistical certainty. Another open question which goes hand in hand with ligand binding: 'does flipping of conserved Tyr/Phe in PAPs necessary during ligand binding to accommodate the ligand in the active site?' Recently, Limongelli[77] reviewed several pathway based simulation methods that can capture conformational change induced ligand binding/unbinding. One of these simulation methods include funnel metadynamics[78] (facilitates frequent ligand binding/unbinding by using a funnel like restraint potential which reduces the sampling of the unbound state) and its variant volume-biased (sphere shaped restraint potential instead of funnel) metadynamics.[79] In theory it is possible to use COM_{unbind} order parameter within funnel/volume biased‡‡ metadynamics framework to understand how conformational changes in PAPs induce ligand binding/unbinding (Fig. 23 and 24). Other enhanced sampling methods such as accelerated MD (aMD),[80–82] weighted ensemble (WE) method,[83–85] adiabatic-biased MD (ABMD)[86,87] *etc* can also be tested on PAPs to explore conformational dynamics, kinetics and free energy associated with ligand binding/unbinding.

‡‡ The restraint potential shouldn't affect the natural dynamics associated with flap region.

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

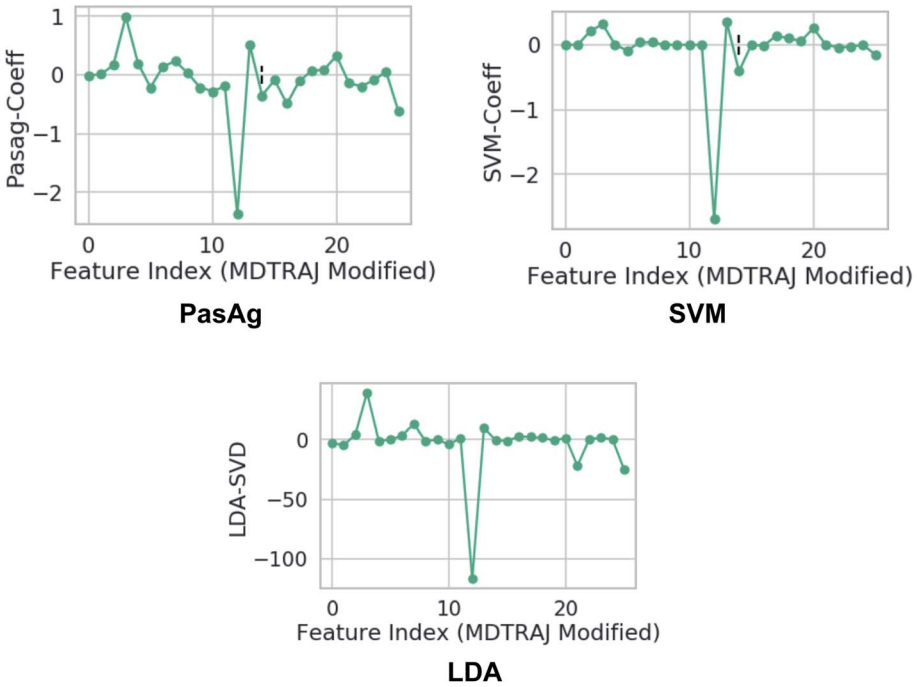

Fig. 18 Weights assign to each features which were used as inputs. Feature 3 and 12 corresponds to $\chi 1$ angle (3 and 12 represents sin $\chi 1$ and cos $\chi 1$ respectively) of Tyr (see Github for more details). One can clearly see that the weights corresponds to $\chi 1$ angle is higher compared to other side-chains angle of the flap region. This gives a clear signal that the $\chi 1$ angle of the conserved Tyr dictates transition between normal and flipped conformations. In case of LDA, the single value decomposition (LDA-SVD) was used as a solver (see scikit-learn for more options).

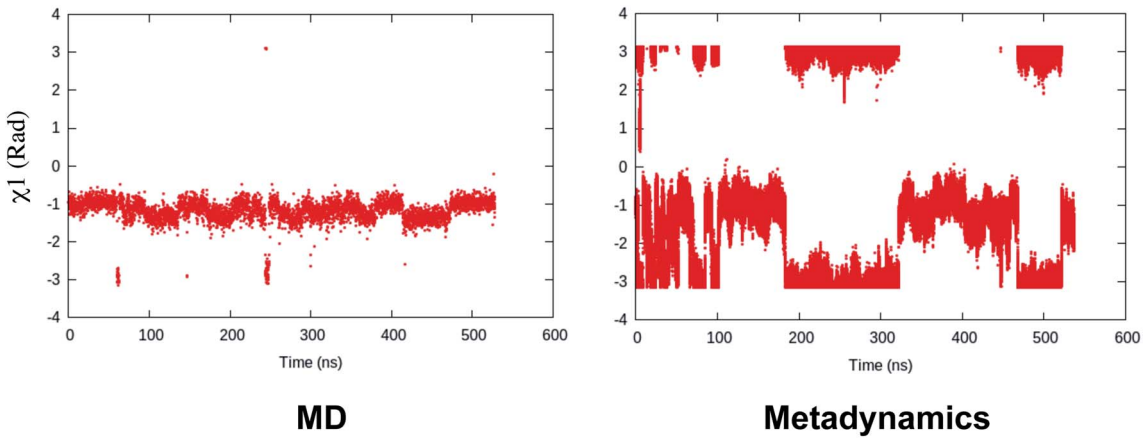

Fig. 19 Fluctuation of $\chi 1$ order parameter during MD simulation of apo PlmII. Metadynamics simulation with PasAg classifier based order parameter shows enhanced fluctuation along $\chi 1$. The plumed input files and tpr files for reproducing this result can be available at Github.

Challenge to the simulation community

Recently, macro-cyclic ligand bound structures of BACE1 were reported by Yen *et al.*[88] (PDB IDs: 6NV7, 6NV9, 6NW3). The authors further reported experimental K_i and K_d values as well as several thermodynamic parameters (ΔH, $T\Delta S$, and ΔG) of ligand binding (thermodynamic parameters were calculated using isothermal titration calorimetry (ITC) experiments). Keeping in mind the availability of experimental datas, size and number of torsion angles of the macro-cyclic ligands and large scale conformational dynamics of BACE1, it will be the perfect challenge for methods *e.g.* funnel metadynamics, volume biased metadynamics, variational autoencoder driven infrequent metadynamics,[89,90] WE,[91] aMD *etc.* To predict binding free energies and kinetics on clinically important protein-ligand systems beyond typical test cases *e.g.* benzamide–trypsin and on systems where the ligand is relatively small with fewer torsional angles and there is no large scale conformational dynamics upon ligand binding.

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

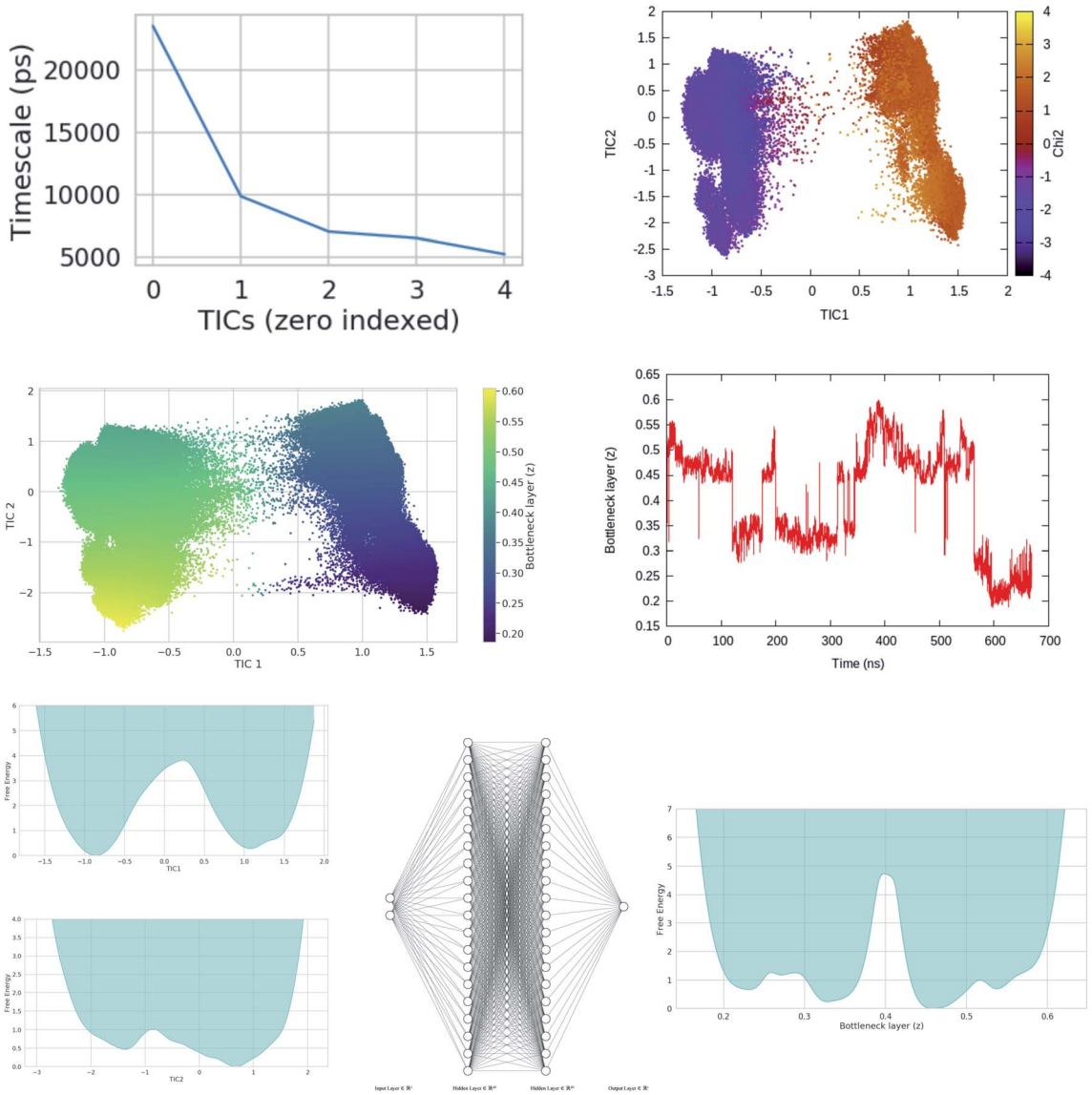

Fig. 20 Upper panel: timescale separation (otherwise known as spectral gap) along different TIC vectors (unbiased MD simulation of apo PlmII with length of ~600 ns). TIC1 and TIC2 projected along $\chi 2$ angle of Tyr-79 shows that $\chi 2$ is one of the slow degrees of freedom in plasmepsin-II (corresponds with the higher weight on sin $\chi 2$ in Fig. 15. Middle panel: bottleneck layer of a variational autoencoder (VAE)[71] projected along TIC1 and TIC2. TIC1–2 were used as inputs in VAE. Different values of bottleneck layer represents different points along TIC vectors. One can see that the bottleneck layer did a better job in separating different metastable states compared to $\chi 2$. Further, time-series projection of the bottleneck layer shows its fluctuation during MD simulation. The coefficients (weights) of the bottleneck layer can be used as order parameters in meta-dynamics simulations. This method in principal similar to combining RAVE[72] with SGOOP.[53] See the Github profile corresponding this article to see step-by-step guide on how to use variational autoencoder. Bottom panel: shows apparent free energy surface projected along TIC1, TIC2 and bottleneck layer. In this case, i have used a variational autoencoder with the following hyper-parameters: number of hidden layers = 2, number of neurons in each hidden layer = 20, number of epochs = 100, batch size = 500, learning rate = $1e-2$.

Combining experiment with molecular simulation

Rotation of Tyr side-chain and opening of the flap region makes PAPs an ideal case to apply methods that combines molecular simulations with biophysical experiments. Combining MD simulation with NMR spin relaxation derived dynamical information[93–97] to capture residue-wise backbone/side-chain configurational entropy is a powerful method to understand stability of proteins and its changes upon perturbation. An analytical approach to capture configurational entropy is to measure residue-wise O^2 (often written as S^2) from MD simulations. It has been shown that configurational entropy of protein can be described from dihedral distribution sampled during MD simulations.[95] For each side-chain dihedral angle ω_j, the probability distribution $p(\omega_j)$ is determined by von-Mises kernel estimation which contributes toward the configurational entropy S:

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

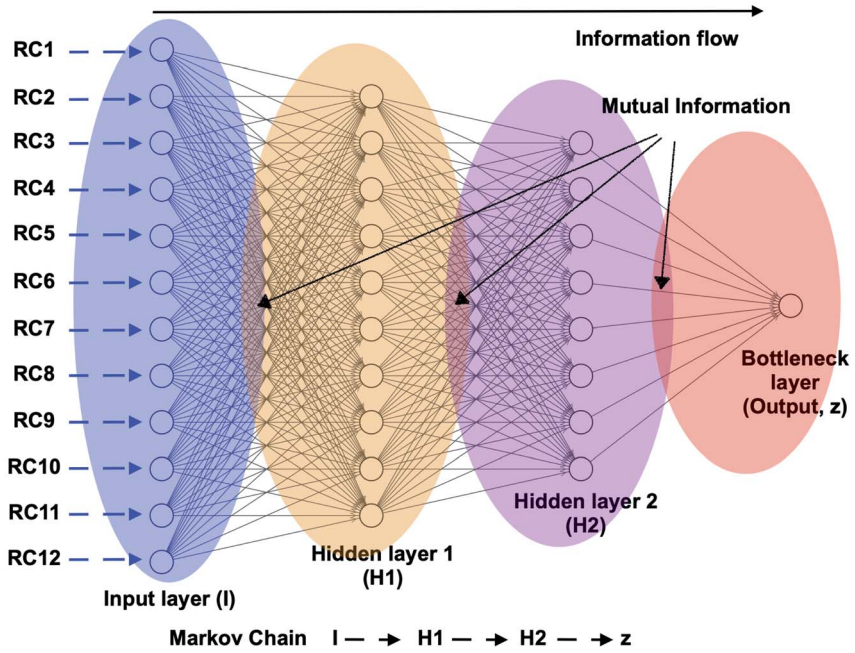

Fig. 21 Schematic representation of a feed forward artificial neural network (ANN) architecture. A typical input to ANN can be multiple RCs which in context of PAPs can be probability distributions of $\chi 1$ and $\chi 2$ angles of the residues in the flap region, tIC/PC/SGOOP vectors sampled during trial MD simulation. The output layer (z) (also known as bottleneck layer) provides a non-linear transformation (sigmoid transformation) of the input features which can be used as a CV in metadynamics simulations. If we think I to z as Markov chains, then it is possible to calculate information flow within ANN by calculating mutual information between each layer (ref. 73 and 74). The hypothesis is that the information output from z will follow principle of maximum entropy (the output will be dominated by input RCs with maximal information entropy).

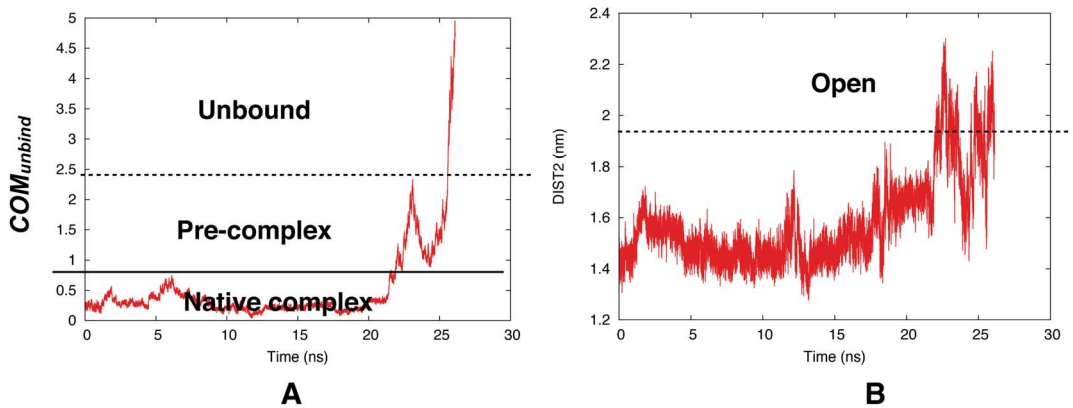

Fig. 22 Time evolution of COM_{unbind} during metadynamics simulations (A). Plotting DIST2 as a function of time shows how flap opening is necessary in order to transit from pre-complex to unbound state. The conserved Tyr remained in *gauche*$^+$ conformation during the unbinding process. The simulation was carried out using PlmII-ligand complex (PDB ID:4Y6M[76]).

$$S = -k_B \sum_{j=1}^{N} \int_0^{2\pi} p_j(\omega_j) \log[p_j(\omega_j)] d\omega_j \quad (3)$$

O^2 is mathematically connected with S using the following equation:

$$S = k_B M[A + Bf(1 - O_{NMR}{}^2)] \quad (4)$$

where M denotes number of side-chain dihedral angles, A and B are fitted parameters. One way to capture O^2 from MD simulation is to use principal component based method otherwise known as isotropic reorientation eigenmode dynamics (iRED).[98] In principal, O^2 derived from MD simulation can be compared with NMR spin relaxation experiments which can give us an idea of residue-wise dynamics associated with the flap region of PAPs (Fig. 25). NMR experiments can further give

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

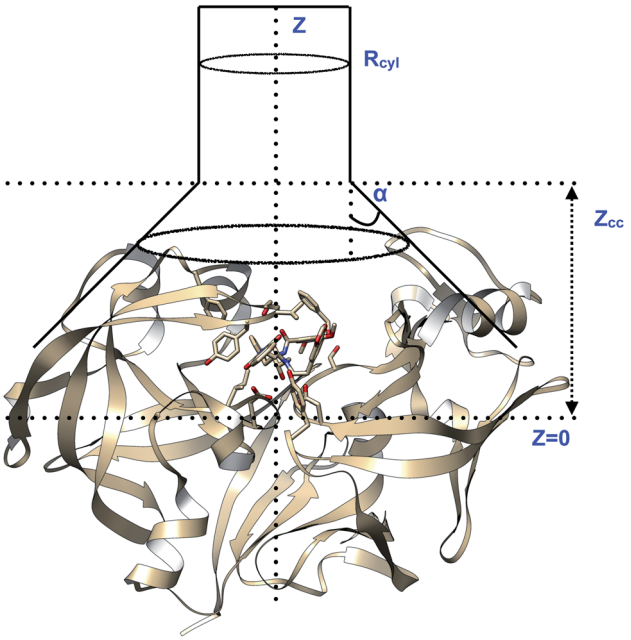

Fig. 23 Pictorial representation of a model funnel shaped potential that can be used to calculate binding free energy and binding/unbinding induced conformational changes in PAPs. R_{cyl} describes the radius of the cylindrical section and Z defines the axis to study binding/unbinding. Z can be COM_{unbind} or some other distance based order parameter. Z_{cc} is the distance where the potential switches from cone to cylindrical shape. The effect of restraint potential can be rigorously calculated as described in ref. 92. R_{cyl} should be adjusted such that it doesn't affect the natural conformational dynamics of the ligand binding/unbinding. PAP-inhibitor complex provides a **true** challenge (compared to T4-lysozyme and other systems where the ligand has few torsional degrees of freedom and protein is relatively **rigid**) for funnel metadynamics as the ligands have large number of torsion angles and the protein undergoes large conformational changes during binding/unbinding. Abstract order parameters *e.g.* TICA/SGOOP/VAC, path CVs can address this challenge. Funnel metadynamics using open flap inhibitors can in principle give insights into how flipping of Tyr and Trp accommodate bigger ligands which has direct application in ligand design.

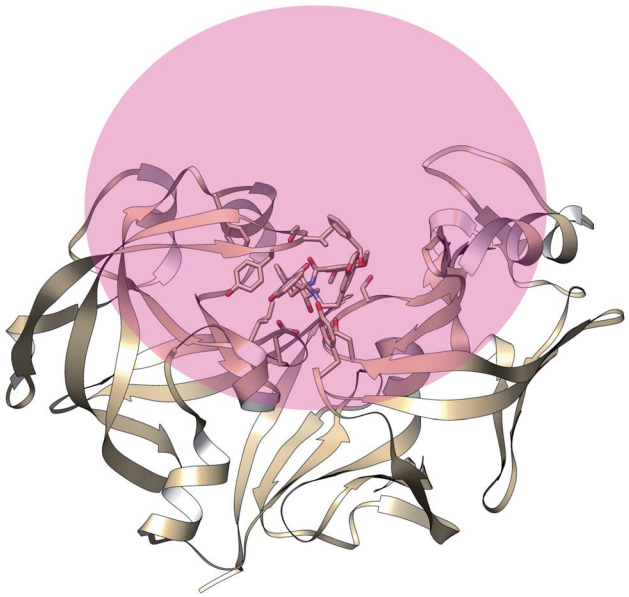

Fig. 24 Pictorial representation of sphere shaped restraint potential applied in case of volume-biased metadynamics. Unlike funnel metadynamics, the sphere shape restraint potential is more general and doesn't assume any prior knowledge of the binding site. This method has been used to identify novel ligand binding pathways in T4-lysozyme. Using this method on PAP-ligand complex can find alternative ligand binding pathways as well as path CVs for binding/unbinding transition. The challenge is to set the radius of the sphere so that it doesn't hamper the large scale conformational dynamics of PAPs upon ligand binding/unbinding.

insights (both dynamics and kinetics) into ring flip associated with conserved Tyr/Phe and Trp in PAPs (ref. 99–101).

Besides combining NMR with MD simulation, time resolved fluorescence spectroscopy, FRET and Raman spectroscopy§§ can be applied to capture conformational dynamics of PAPs. This is an open area of research which didn't gain much attention from the biophysical or structural biology community. In theory, time-dependent order parameters from experiments (*e.g.* FRET) can be integrated with MD simulations using embedding theorem *e.g.* Taken's embedding theorem (ref. 104). Ligand binding to PAPs can be calculated using isothermal titration calorimetry (ITC)[105] or differential scanning calorimetry (DSC)[106] which measures different thermodynamic parameters involved in the binding process. The major challenge in front of the simulation community is to calculate thermodynamic parameters from pathway based simulation methods (or from enhanced sampling calculations) that agrees with experiment. This is a growing area of research within biophysical community and PAPs can be an ideal test system for integrating biophysical experiments with biomolecular simulation.

§§ Raman spectra can provide H-bonding pattern of Tyr and Trp side-chain residues. If Tyr residue shows Fermi resonance in Raman spectra that gives an indication of the rotational degrees of freedom associated with the Tyr side-chain.[103]

Ideal playing field for drug repurposing

Plasmepsins (especially PlmII, PlmIV, PlmV, PlmIX and X) are drug targets for malaria.[107,108] Artemisinin and chloroquine are the two major blockbuster drugs against malaria. Emergence of artemisinin[109] and chloroquine[110] resistant *P. falciparum* possess a huge risk towards eradication efforts towards malaria in Africa and South-East Asia. This calls for identification of novel drug targets against malaria and plasmepsins has potential to become the next big target against malaria. BACE-1 is a promising drug target for Alzheimer's disease with several inhibitors in different stages clinical trials and several ligand bound crystal structures deposited in PDB. Due to structural similarity and similar mechanism of action which governs conformational dynamics of apo PlmII and BACE-1 (in general all PAPs), the ligands targeting BACE-1 can be repurposed on plasmepsins. Machine learning based methods (*e.g.* convolutional neural network) combined with molecular

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

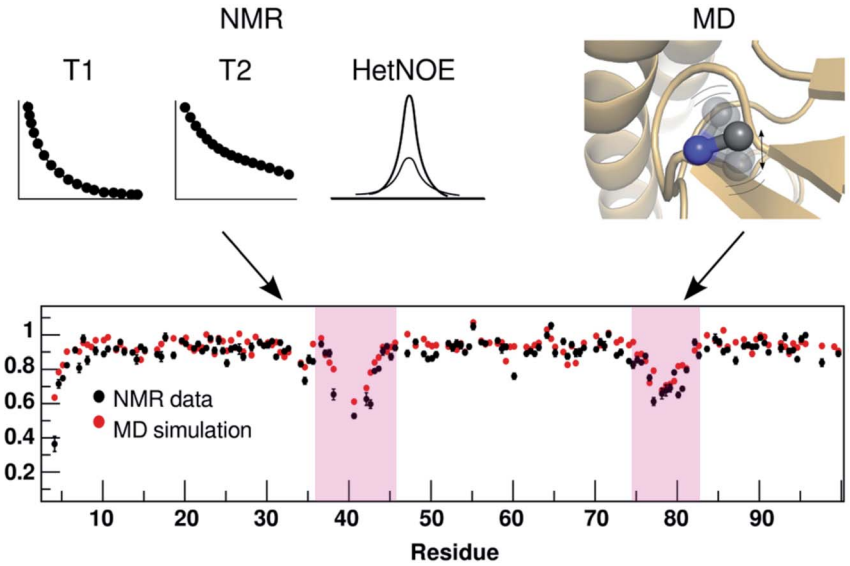

Fig. 25 Comparison of residue-wise dynamical properties measured by NMR and MD derived order parameter O^2 (y axis in the bottom panel). The backbone order parameter provides information on the amplitude of fluctuation of NH and CH bond vectors. The magenta shadowed area shows the residues with higher backbone flexibility (lower O^2 corresponds to higher flexibility). Experimental parameters can be calculated by T1, T2 and heteronuclear NOE relaxation experiments. Often short MD simulations are sufficient to produce O^2 which can capture motions in ps to ns timescale but for slower relaxation motions one needs longer simulations (μs to ms). Comparison of NMR and MD derived O^2 in PAPs can give insights into how flap flexibility varies between apo and ligand-bound conformations. See ref. 95 and 96 for comparison of conformational dynamics using MD and NMR. The figure was adapted from ref. 102.

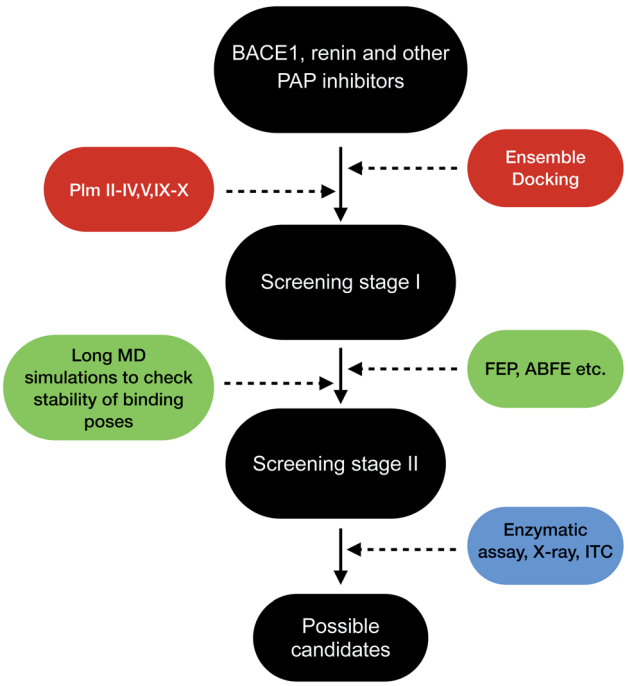

Fig. 26 Schematic diagram showing a workflow on how computational predictions and experiments can work coherently in order to repurpose BACE1 and other PAP inhibitors on plasmepsins. Growing interest of artificial intelligence (AI) based drug discovery companies (*e.g.* Atomwise) within drug repurposing in the wake of COVID19 shows that AI can also be used within drug repurposing workflow in context of plasmepsins.

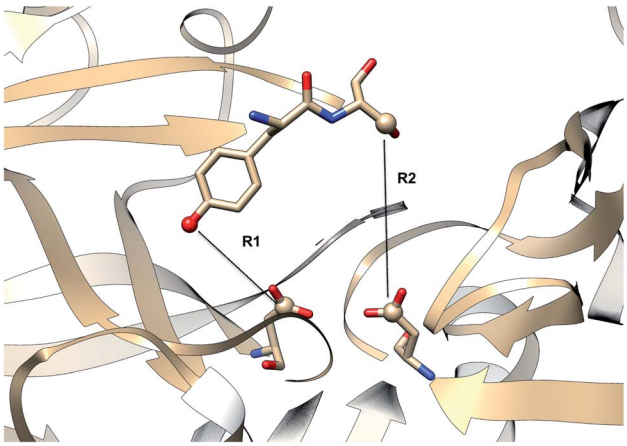

Fig. 27 Distance parameters proposed by Shen and co-workers. R1: distance between Tyr-83-OH and Asp-38-CG and R2: Ser-84-CB and Asp-38-CG. The figure was created using PDB:2REN.

docking,[111–113] free energy calculations[114] (*e.g.* FEP, ABFE, MM/PBSA or MM/GBSA) and biochemical assay can then be used to identify BACE-1 inhibitors (in broad sense this concept is applicable for other PAP inhibitors *e.g.* human renin inhibitor) as potential plasmepsin inhibitors. This repurposing strategy can help to identify scaffolds¶¶ that can be further modified by synthetic chemists to develop potent molecules targeting

¶¶ Without designing ligands from scratch.

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

plasmepsins (Fig. 26). I believe in future machine learning driven molecular docking combined with experiment will play an important role in repurposing other PAP inhibitors against plasmepsins.

Conclusions and future directions

The main aim of this review was to highlight the possibilities of using PAPs as model systems to test novel simulation methods and to combine biomolecular simulation with biophysical experiments. In order to kick-start this effort I have curated a Github profile which contains inputs for performing biomolecular simulation and order parameters that can be monitored during molecular simulation. Although in practice there has been a few advances in application of state-of-the art biomolecular simulations on PAPs, a combination of biophysical experiments with biomolecular simulation (otherwise known as integrative structural biology[115]) is far from a standard practice. Dynamic nature of PAPs makes it an ideal case to integrate experiment with molecular simulation using methods such as Bayesian-maximum entropy[116] *etc.* which can provide both atomic-level description and mechanistic understanding (*e.g.* population of normal/flipped states, role of force field and water models in population of different states, role of Tyr/Phe in conformational dynamics and ligand binding, extent of flap opening *etc.*) of the PAPs.

Finally, COVID-19 pandemic shows a resurgence of drug repurposing by combining molecular simulation with biochemical experiments. Projects like COVID Moonshot shows how an old simulation technique (free energy perturbation) can be effectively used in order to hunt novel drug candidates against COVID. Similar simulation strategies (AI driven molecular docking, free energy perturbation *etc.*) can be applied to re-purpose BACE1, human renin, pepsin inhibitors against plasmepsins followed up by experimental validation (biochemical assay, X-ray crystallography, NMR *etc.*). Further keeping in mind the neglected[117] status of malaria it is of utmost importance to share computational and experimental protocols openly using public data/code-sharing platforms *e.g.* Jupyter Notebook, Github, Google Collab *etc.* Finally, bigger and dynamical nature PAPs and their role in human disease makes it an ideal model system for experimentalists and computational chemists to work together towards developing new methods and applying old methods which will have direct application towards developing novel therapeutics against malaria, Alzheimer's disease *etc.*

Note

To date, no pathway based simulation methods able to predict thermodynamics and kinetic parameters of HIV protease inhibitors (for the sake of sample size we can focus on nine clinically approved drugs) and compare it with experiments. Even in the year 2020, the most widely used method for small molecule drug discovery (clinically relevant) is molecular docking (often followed by visual inspection and FEP) which is computationally less expensive compared to physical pathway based methods. Two major problems of molecular docking are approximated energy functions and lack of sampling (especially of the protein). The remedy to this is a skin-in-the game problem.

'when the solution is about solving this very problem-Nassim Nicholas Taleb'.

Observation

In D3R Grand Challenge 4 (ref. 118) which aims at predicting binding pose and free energy using BACE1, the top performing submissions didn't use pathway based sampling methods.

Conflicts of interest

Author declares no potential conflicts of interest.

Acknowledgements

SB thanks Swedish A-kassa for the financial support. The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at LUNARC (Lund University) and HPC2N (Umeå University).

References

1 B. M. Dunn, Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases, *Chem. Rev.*, 2002, **102**, 4431–4458.
2 N. D. Rawlings, A. J. Barrett, P. D. Thomas, X. Huang, A. Bateman and R. D. Finn, The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database, *Nucleic Acids Res.*, 2017, **46**, D624–D632.
3 M. Mahanti, S. Bhakat, U. J. Nilsson and P. Söderhjelm, Flap Dynamics in Aspartic Proteases: A Computational Perspective, *Chem. Biol. Drug Des.*, 2016, **88**, 159–177.
4 N. M. Moussa-Pacha, S. M. Abdin, H. A. Omar, H. Alniss and T. H. Al-Tel, BACE1 inhibitors: Current status and future directions in treating Alzheimer's disease, *Med. Res. Rev.*, 2020, **40**, 339–384.
5 W. Karubiu, S. Bhakat, L. McGillewie and M. E. S. Soliman, Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics, *Mol. BioSyst.*, 2015, **11**, 1061–1066.
6 S. Bhakat and P. Söderhjelm, Flap dynamics in pepsin-like aspartic proteases: a computational perspective using Plasmepsin-II and BACE-1 as model systems, *bioRxiv*, 2020, 1–47.
7 H. M. Kumalo, S. Bhakat and M. E. Soliman, Investigation of flap flexibility of β-secretase using molecular dynamic simulations, *J. Biomol. Struct. Dyn.*, 2016, **34**, 1008–1019.
8 H. M. Kumalo and M. E. Soliman, A comparative molecular dynamics study on BACE1 and BACE2 flap flexibility, *J. Recept. Signal Transduction*, 2016, **36**, 505–514.

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

9 A. A. Gorfe and A. Caflisch, Functional Plasticity in the Substrate Binding Site of β−Secretase, *Structure*, 2005, **13**, 1487–1498.
10 S. A. Spronk and H. A. Carlson, The role of tyrosine 71 in modulating the flap conformations of BACE1, *Proteins: Struct., Funct., Bioinf.*, 2011, **79**, 2247–2259.
11 S. Bottaro and K. Lindorff-Larsen, Biophysical experiments and biomolecular simulations: A perfect match?, *Science*, 2018, **361**, 355–360.
12 Y. Xu, M.-j. Li, H. Greenblatt, W. Chen, A. Paz, O. Dym, Y. Peleg, T. Chen, X. Shen, J. He, H. Jiang, I. Silman and J. L. Sussman, Flexibility of the flap in the active site of BACE1 as revealed by crystal structures and molecular dynamics simulations, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 13–25.
13 O. A. Asojo, S. V. Gulnik, E. Afonina, B. Yu, J. A. Ellman, T. S. Haque and A. M. Silva, Novel Uncomplexed and Complexed Structures of Plasmepsin II, an Aspartic Protease from Plasmodium falciparum, *J. Mol. Biol.*, 2003, **327**, 173–181.
14 A. N. Hodder, B. E. Sleebs, P. E. Czabotar, M. Gazdik, Y. Xu, M. T. O'Neill, S. Lopaticki, T. Nebl, T. Triglia, B. J. Smith, K. Lowes, J. A. Boddey and A. F. Cowman, Structural basis for plasmepsin V inhibition that blocks export of malaria proteins to human erythrocytes, *Nat. Struct. Mol. Biol.*, 2015, **22**, 590–596.
15 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Res.*, 2018, **46**, W296–W303.
16 L. Prade, A. F. Jones, C. Boss, S. Richard-Bildstein, S. Meyer, C. Binkert and D. Bur, X-ray Structure of Plasmepsin II Complexed with a Potent Achiral Inhibitor, *J. Biol. Chem.*, 2005, **280**, 23837–23843.
17 C. Boss, O. Corminboeuf, C. Grisostomi, S. Meyer, A. Jones, L. Prade, C. Binkert, W. Fischli, T. Weller and D. Bur, Achiral, Cheap, and Potent Inhibitors of Plasmepsins I, II, and IV, *ChemMedChem*, 2006, **1**, 1341–1345.
18 D. Rasina, M. Otikovs, J. Leitans, R. Recacha, O. V. Borysov, I. Kanepe-Lapsa, I. Domraceva, T. Pantelejevs, K. Tars, M. J. Blackman, K. Jaudzems and A. Jirgensons, Fragment-Based Discovery of 2-Aminoquinazolin-4(3H)-ones As Novel Class Nonpeptidomimetic Inhibitors of the Plasmepsins I, II, and IV, *J. Med. Chem.*, 2016, **59**, 374–387, PMID: 26670264.
19 A. Y. Lee, S. V. Gulnik and J. W. Erickson, Conformational switching in an aspartic proteinase, *Nat. Struct. Biol.*, 1998, **5**, 866–871.
20 E. T. Baldwin, T. N. Bhat, S. Gulnik, M. V. Hosur, R. C. Sowder, R. E. Cachau, J. Collins, A. M. Silva and J. W. Erickson, Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 6796–6800.
21 V. Hornak, A. Okur, R. C. Rizzo and C. Simmerling, HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state, *J. Am. Chem. Soc.*, 2006, **128**, 2812–2813, Using Smart Source Parsing Mar 8.
22 V. Hornak, A. Okur, R. C. Rizzo and C. Simmerling, HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 915–920.
23 S. Ma, J. A. Henderson and J. Shen, Exploring the pH-dependent structure-dynamics-function relationship of human renin, *J. Chem. Inf. Model.*, 2021, **61**(1), 400–407.
24 S. Patel, L. Vuillard, A. Cleasby, C. W. Murray and J. Yon, Apo and Inhibitor Complex Structures of BACE (β − *secretase*), *J. Mol. Biol.*, 2004, **343**, 407–416.
25 L. Hong and J. Tang, Flap Position of Free Memapsin 2 (β-Secretase), a Model for Flap Opening in Aspartic Protease Catalysis, *Biochemistry*, 2004, **43**, 4689–4695.
26 C. Oefner, *et al.*, Renin inhibition by substituted piperidines: A novel paradigm for the inhibition of monomeric aspartic proteinases?, *Chemistry and Biology*, 1999, **6**, 127–131.
27 R. Bobrovs, K. Jaudzems and A. Jirgensons, Exploiting Structural Dynamics To Design Open-Flap Inhibitors of Malarial Aspartic Proteases, *J. Med. Chem.*, 2019, **62**, 8931–8950.
28 G. L. Gilliland, E. L. Winborne, J. Nachman and A. Wlodawer, The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution, *Proteins: Struct., Funct., Bioinf.*, 1990, **8**, 82–101.
29 F. Suzuki, K. Goto, Y. Shiratori, T. Inagami, K. Murakami and Y. Nakamura, Tyrosine 83 of renin has an important role in renin?angiotensinogen reaction, *Protein Pept. Lett.*, 1996, **3**, 45–49.
30 Y.-N. Park, J.-i. Aikawa, M. Nishiyama, S. Horinouchi and T. Beppu, Involvement of a residue at position 75 in the catalytic mechanism of a fungal aspartic proteinase, Rhizomucor pusilus pepsin. Replacement of tyrosine 75 on the flap by asparagine enhances catalytic efficiency, *Protein Eng., Des. Sel.*, 1996, **9**, 869–875.
31 F. Sittel, A. Jain and G. Stock, Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates, *J. Chem. Phys.*, 2014, **141**, 014111.
32 C. C. David, and D. J. Jacobs, in *Protein Dynamics: Methods and Protocols*, ed. D. R. Livesay, Humana Press, Totowa, NJ, 2014, pp. 193–226.
33 S. A. M. Stein, A. E. Loccisano, S. M. Firestine, and J. D. Evanseck, in Chapter 13 Principal Components Analysis: A Review of its Application on Molecular Dynamics Data, *Annual Reports in Computational Chemistry*, ed. D. C. Spellmeyer, Elsevier, 2006; Vol. 2, pp 233 – 261.
34 Y. Naritomi and S. Fuchigami, Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis, *J. Chem. Phys.*, 2013, **139**, 215102.

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

35 T. Blaschke, P. Berkes and L. Wiskott, What Is the Relation Between Slow Feature Analysis and Independent Component Analysis?, *Neural Comput.*, 2006, **18**, 2495–2508.
36 L. Molgedey and H. G. Schuster, Separation of a mixture of independent signals using time delayed correlations, *Phys. Rev. Lett.*, 1994, **72**, 3634–3637.
37 C. R. Schwantes and V. S. Pande, Modeling Molecular Kinetics with tICA and the Kernel Trick, *J. Chem. Theory Comput.*, 2015, **11**, 600–608.
38 J. Shlens, *A Tutorial on Independent Component Analysis*, CoRR abs/1404.2986, 2014.
39 H. Zhou, F. Wang and P. Tao, t-Distributed Stochastic Neighbor Embedding Method with the Least Information Loss for Macromolecular Simulations, *J. Chem. Theory Comput.*, 2018, **14**, 5499–5510.
40 V. Spiwok and P. Kříž, Time-Lagged t-Distributed Stochastic Neighbor Embedding (t-SNE) of Molecular Simulation Trajectories, *Front. Mol. Biosci.*, 2020, **7**, 132.
41 M. Wattenberg, F. Viégas and I. Johnson, How to Use t-SNE Effectively, *Distill*, 2016, DOI: 10.23915/distill.00002.
42 L. van der Maaten and G. Hinton, Visualizing Data using t-SNE, *J. Mach. Learn. Res.*, 2008, **9**, 2579–2605.
43 A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis and P. G. Debenedetti, Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach, *Chem. Phys. Lett.*, 2011, **509**, 1–11.
44 W. Zheng, M. A. Rohrdanz and C. Clementi, Rapid Exploration of Configuration Space with Diffusion-Map-Directed Molecular Dynamics, *J. Phys. Chem. B*, 2013, **117**, 12769–12776.
45 M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess and E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1–2**, 19–25.
46 M. P. Harrigan, M. M. Sultan, C. X. Hernández, B. E. Husic, P. Eastman, C. R. Schwantes, K. A. Beauchamp, R. T. McGibbon and V. S. Pande, MSMBuilder: Statistical Models for Biomolecular Dynamics, *Biophys. J.*, 2017, **112**, 10–15.
47 R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane and V. S. Pande, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories, *Biophys. J.*, 2015, **109**, 1528–1532.
48 M. J. McKeown, L. K. Hansen and T. J. Sejnowsk, Independent component analysis of functional MRI: what is signal and what is noise?, *Curr. Opin. Neurobiol.*, 2003, **13**, 620–629.
49 L. Sun,, Y. Liu,, and P. J. Beadle, Independent component analysis of EEG signals, *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology*, 2005, pp. 219–222.
50 S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte and F. Noé, Data-Driven Model Reduction and Transfer Operator Approximation, *J. Nonlinear Sci.*, 2018, **28**, 985–1010.
51 M. Sultan and V. S. Pande, tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables, *J. Chem. Theory Comput.*, 2017, **13**, 2440–2447.
52 J. McCarty and M. Parrinello, A variational conformational dynamics approach to the selection of collective variables in metadynamics, *J. Chem. Phys.*, 2017, **147**, 204109.
53 P. Tiwary and B. J. Berne, Spectral gap optimization of order parameters for sampling complex molecular systems, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 2839–2844.
54 K. Ghosh, P. D. Dixit, L. Agozzino and K. A. Dill, The Maximum Caliber Variational Principle for Nonequilibria, *Annu. Rev. Phys. Chem.*, 2020, **71**, 213–238, PMID: 32075515.
55 A. J. Bell and T. J. Sejnowski, An Information-Maximization Approach to Blind Separation and Blind Deconvolution, *Neural Comput.*, 1995, **7**, 1129–1159.
56 A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley and Sons, Ltd, 2001, ch. 10, pp. 221–227.
57 D. Rutledge and D. Jouan-Rimbaud Bouveresse, Independent Components Analysis with the JADE algorithm, *TrAC, Trends Anal. Chem.*, 2013, **50**, 22–32.
58 A. Hyvärinen and E. Oja, Independent component analysis: algorithms and applications, *Neural Network.*, 2000, **13**, 411–430.
59 M. M. Sultan and V. S. Pande, Automated design of collective variables using supervised machine learning, *J. Chem. Phys.*, 2018, **149**, 094106.
60 C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.*, 1995, **20**, 273–297.
61 K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz and Y. Singer, Online Passive-Aggressive Algorithms, *J. Mach. Learn. Res.*, 2006, **7**, 551–585.
62 H.-F. Yu, F.-L. Huang and C.-J. Lin, Dual coordinate descent methods for logistic regression and maximum entropy models, *Mach. Learn.*, 2011, **85**, 41–75.
63 Y. Freund and R. E. Schapire, Large Margin Classification Using the Perceptron Algorithm, *Mach. Learn.*, 1999, **37**, 277–296.
64 A. M. Martinez and A. C. Kak, PCA versus LDA, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, 228–233.
65 D. Mendels, G. Piccini and M. Parrinello, Collective Variables from Local Fluctuations, *J. Phys. Chem. Lett.*, 2018, **9**, 2776–2781.
66 H. Sidky, W. Chen and A. L. Ferguson, Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation, *Mol. Phys.*, 2020, **118**, e1737742.
67 Y. Wang, J. M. Lamim Ribeiro and P. Tiwary, Machine learning approaches for analyzing and enhancing molecular dynamics simulations, *Curr. Opin. Struct. Biol.*, 2020, **61**, 139–145, Theory and Simulation: Macromolecular Assemblies.
68 F. Creutzig and H. Sprekeler, Predictive Coding and the Slowness Principle: An Information-Theoretic Approach, *Neural Comput.*, 2008, **20**, 1026–1041.

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

69 S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Comput.*, 1997, **9**, 1735–1780.
70 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin *Attention Is All You Need*. 2017.
71 C. X. Hernández, H. K. Wayment-Steele, M. M. Sultan, B. E. Husic and V. S. Pande, Variational encoding of complex dynamics, *Phys. Rev. E*, 2018, **97**, 062412.
72 Y. Wang, J. M. L. Ribeiro and P. Tiwary, Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics, *Nat. Commun.*, 2019, **10**, 3573.
73 Z. Goldfeld, and Y. Polyanskiy *The Information Bottleneck Problem and Its Applications in Machine Learning*. 2020.
74 R. Shwartz-Ziv, and N. Tishby *Opening the Black Box of Deep Neural Networks via Information*. 2017.
75 Y. Yu, J. Wang, Z. Chen, G. Wang, Q. Shao, J. Shi and W. Zhu, Structural insights into HIV-1 protease flap opening processes and key intermediates, *RSC Adv.*, 2017, **7**, 45121–45128.
76 R. Recacha, J. Leitans, I. Akopjana, L. Aprupe, P. Trapencieris, K. Jaudzems, A. Jirgensons and K. Tars, Structures of plasmepsin II from *Plasmodium falciparum* in complex with two hydroxyethylamine-based inhibitors, *Acta Crystallogr., Sect. F: Struct. Biol. Commun.*, 2015, **71**, 1531–1539.
77 V. Limongelli, Ligand binding free energy and kinetics calculation in 2020, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2020, **10**, e1455.
78 S. Raniolo and V. Limongelli, Ligand binding free-energy calculations with funnel metadynamics, *Nat. Protoc.*, 2020, **15**, 2837–2866.
79 R. Capelli, P. Carloni and M. Parrinello, Exhaustive Search of Ligand Binding Pathways via Volume-Based Metadynamics, *J. Phys. Chem. Lett.*, 2019, **10**, 3495–3499.
80 K. Kappel, Y. Miao and J. A. McCammon, Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor, *Q. Rev. Biophys.*, 2015, **48**, 479–487.
81 M. Araki, S. Matsumoto, G.-J. Bekker, Y. Isaka, Y. Sagae, N. Kamiya and Y. Okuno, Exploring ligand binding pathways on proteins using hypersound–accelerated molecular dynamics, *bioRxiv*, 2020, 1–14.
82 D. B. Kokh, M. Amaral, J. Bomke, U. Grädler, D. Musil, H.-P. Buchstaller, M. K. Dreyer, M. Frech, M. Lowinski, F. Vallee, M. Bianciotto, A. Rak and R. C. Wade, Estimation of Drug-Target Residence Times by τ−Random Acceleration Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 2018, **14**, 3859–3869.
83 A. Nunes-Alves, D. M. Zuckerman and G. M. Arantes, Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways, *Biophys. J.*, 2018, **114**, 1058–1066.
84 S.-H. Ahn, B. Jagger and R. E. Amaro, Ranking of Ligand Binding Kinetics using a Weighted Ensemble Approach and Comparison with Milestoning, *Biophys. J.*, 2020, **118**, 305a.
85 A. Dickson and S. D. Lotz, Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore, *Biophys. J.*, 2017, **112**, 620–629.
86 J. M. Johnston, and M. Filizola in *G Protein-Coupled Receptors - Modeling and Simulation*, ed.M. Filizola, Springer Netherlands, Dordrecht, 2014, pp. 95–125.
87 M. Marchi and P. Ballone, Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems, *J. Chem. Phys.*, 1999, **110**, 3697–3702.
88 Y.-C. Yen, A. M. Kammeyer, K. C. Jensen, J. Tirlangi, A. K. Ghosh and A. D. Mesecar, Development of an Efficient Enzyme Production and Structure-Based Discovery Platform for BACE1 Inhibitors, *Biochemistry*, 2019, **58**, 4424–4435.
89 R. Casasnovas, V. Limongelli, P. Tiwary, P. Carloni and M. Parrinello, Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations, *J. Am. Chem. Soc.*, 2017, **139**, 4780–4788.
90 J. M. Lamim Ribeiro, D. Provasi and M. Filizola, A combination of machine learning and infrequent metadynamics to efficiently predict kinetic rates, transition states, and molecular determinants of drug dissociation from G protein-coupled receptors, *J. Chem. Phys.*, 2020, **153**, 124105.
91 A. Dickson and S. D. Lotz, Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore, *Biophys. J.*, 2017, **112**, 620–629.
92 V. Limongelli, M. Bonomi and M. Parrinello, Funnel metadynamics as accurate binding free-energy method, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 6358–6363.
93 N. Trbovic, B. Kim, R. A. Friesner and A. G. Palmer III, Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation, *Proteins: Struct., Funct., Bioinf.*, 2008, **71**, 684–694.
94 K. A. Sharp, E. O'Brien, V. Kasinath and A. J. Wand, On the relationship between NMR-derived amide order parameters and protein backbone entropy changes, *Proteins: Struct., Funct., Bioinf.*, 2015, **83**, 922–930.
95 D.-W. Li and R. Brüschweiler, A Dictionary for Protein Side-Chain Entropies from NMR Order Parameters, *J. Am. Chem. Soc.*, 2009, **131**, 7226–7227.
96 Y. Gu, D.-W. Li and R. Brüschweiler, NMR Order Parameter Determination from Long Molecular Dynamics Trajectories for Objective Comparison with Experiment, *J. Chem. Theory Comput.*, 2014, **10**, 2599–2607.
97 A. Villa and S. Gerhard, What NMR Relaxation Can Tell Us about the Internal Motion of an RNA Hairpin: A Molecular Dynamics Simulation Study, *J. Chem. Theory Comput.*, 2006, **2**, 1228–1236, DOI: 10.1021/ct600160z.
98 J. J. Prompers and R. Brüschweiler, General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation, *J. Am. Chem. Soc.*, 2002, **124**, 4522–4534.
99 U. Weininger, K. Modig and M. Akke, Ring Flips Revisited: 13C Relaxation Dispersion Measurements of Aromatic Side

© 2021 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 17 March 2021. Downloaded on 2/25/2026 12:58:36 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Chain Dynamics and Activation Barriers in Basic Pancreatic Trypsin Inhibitor, *Biochemistry*, 2014, **53**, 4519–4525.

100 M. Dreydoppel, H. N. Raum and U. Weininger, Slow ring flips in aromatic cluster of GB1 studied by aromatic 13C relaxation dispersion methods, *J. Biomol. NMR*, 2020, **74**, 183–191.

101 M. Hattori, H. Li, H. Yamada, K. Akasaka, W. Hengstenberg, W. Gronwald and H. R. Kalbitzer, Infrequent cavity-forming fluctuations in HPr from Staphylococcus carnosus revealed by pressure and temperature dependent tyrosine ring flips, *Protein Sci.*, 2004, **13**, 3104–3114.

102 N. Pastor and C. Amero, Information flow and protein dynamics: the interplay between nuclear magnetic resonance spectroscopy and molecular dynamics simulations, *Front. Plant Sci.*, 2015, **6**, 306.

103 M. Valldeperas, M. Talaikis, S. K. Dhayal, M. Velička, J. Barauskas, G. Niaura and T. Nylander, Encapsulation of Aspartic Protease in Nonlamellar Lipid Liquid Crystalline Phases, *Biophys. J.*, 2019, **117**, 829–843.

104 J. Wang and A. L. Ferguson, Nonlinear reconstruction of single-molecule free-energy surfaces from univariate time series, *Phys. Rev. E*, 2016, **93**, 032412.

105 H. Su and Y. Xu, Application of ITC-Based Characterization of Thermodynamic and Kinetic Association of Ligands With Proteins in Drug Design, *Front. Pharmacol.*, 2018, **9**, 1133.

106 M. S. Celej, S. A. Dassie, M. Gonzalez, M. L. Bianconi and G. D. Fidelio, Differential scanning calorimetry as a tool to estimate binding parameters in multiligand binding proteins, *Anal. Biochem.*, 2006, **350**, 277–284.

107 N. Dan and S. Bhakat, New paradigm of an old target: An update on structural biology and current progress in drug design towards plasmepsin {II}, *Eur. J. Med. Chem.*, 2015, **95**, 324–348.

108 P. M. Cheuka, G. Dziwornu, J. Okombo and K. Chibale, Plasmepsin Inhibitors in Antimalarial Drug Discovery: Medicinal Chemistry and Target Validation (2000 to Present), *J. Med. Chem.*, 2020, **63**, 4445–4467.

109 A. M. Dondorp, *et al.*, Artemisinin Resistance in Plasmodium falciparum Malaria, *N. Engl. J. Med.*, 2009, **361**, 455–467, PMID: 19641202.

110 T. E. Wellems and C. V. Plowe, Chloroquine-Resistant Malaria, *J. Infect. Dis.*, 2001, **184**, 770–776.

111 H. Jiang, M. Fan, J. Wang, A. Sarma, M. Shruti, V. Dokholyan Nikolay, M. Mehrdad and T. Kandemir Mahmut, Guiding Conventional Protein–Ligand Docking Software with Convolutional Neural Networks, *J. Chem. Inf. Model.*, 2020, **60**, 4594–4602, DOI: 10.1021/acs.jcim.0c00542.

112 B. Wang and Ng Ho-Leung, Deep neural network affinity model for BACE inhibitors in D3R Grand Challenge 4, *J. Comput.-Aided Mol. Des.*, 2020, **34**, 201–217.

113 I. Wallach, M. Dzamba, and A. Heifets, *AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery*. CoRR abs/1510.02855, 2015.

114 C. Zoe, B. Allen and W. Sherman, Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations, *J. Chem. Inf. Model.*, 2017, **57**, 2911–2937, DOI: 10.1021/acs.jcim.7b00564.

115 M. L. Verteramo, O. Stenström, M. M. Ignjatović, O. Caldararu, M. A. Olsson, F. Manzoni, H. Leffler, E. Oksanen, D. T. Logan, U. J. Nilsson, U. Ryde and M. Akke, Interplay between Conformational Entropy and Solvation Entropy in Protein–Ligand Binding, *J. Am. Chem. Soc.*, 2019, **141**, 2012–2026.

116 J. He and K. Alexander, Bayesian maximum entropy approach and its applications: a review, *Stoch. Environ. Res. Risk Assess.*, 2018, **32**, 859–877.

117 M. Allarakhia and L. Ajuwon, Understanding and creating value from open source drug discovery for neglected tropical diseases, *Expet Opin. Drug Discov.*, 2012, **7**, 643–657, PMID: 22657529.

118 D. Parks Conor, Z. Gaieb, M. Chiu, H. Yang, C. Shao, W. W. Patrick, M. Jansen Johanna, McG. Georgia, A. Lewis Richard, D. Bembenek Scott, K. Ameriks Michael, M. Tara, K. Burley Stephen, E. Amaro Rommie and K. Gilson Michael, D3R grand challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies, *J. Comput.-Aided Mol. Des.*, 2020, **34**, 99–119.

© 2021 The Author(s). Published by the Royal Society of Chemistry