

# PBmapclust: Mapping and Clustering the Protein Conformational Space Using a Structural Alphabet

I. Vetrivel<sup>1</sup>, L. Hoffmann<sup>2</sup>, S. Guegan<sup>1</sup>, B. Offmann<sup>2</sup> and A. D. Laurent<sup>1</sup>

<sup>1</sup>Université de Nantes, CEISAM UMR CNRS 6230, 2 chemin de la Houssiniere, 44322 Nantes, France

<sup>2</sup>Université de Nantes, UFIP UMR CNRS 6286, 2 chemin de la Houssiniere, 44322 Nantes, France

## Abstract

Analyzing the data from molecular dynamics simulation of biological macromolecules like proteins is challenging. We propose a simple tool called PBmapclust that is based on a well established structural alphabet called Protein blocks (PB). PBs help in tracing the trajectory of the protein backbone by categorizing it into 16 distinct structural states. PBmapclust provides a time vs. amino acid residue plot that is color coded to match each of the PBs. Color changes correspond to structural changes, giving a visual overview of the simulation. Further, PBmapclust enables the user to "map" the conformational space sampled by the protein during the MD simulation by clustering the conformations. The ability to generate sub-maps for specific residues and specific time intervals allows the user to focus on residues of interest like for active sites or disordered regions. We have included an illustrative case study to demonstrate the utility of the tool. It describes the effect of the disordered domain of a HSP90 co-chaperone on the conformation of its active site residues. The scripts required to perform PBmapclust are made freely available under the GNU general public license.

## CCS Concepts

•Applied computing → Molecular structural biology; Bioinformatics; Computational proteomics;

## 1. Introduction

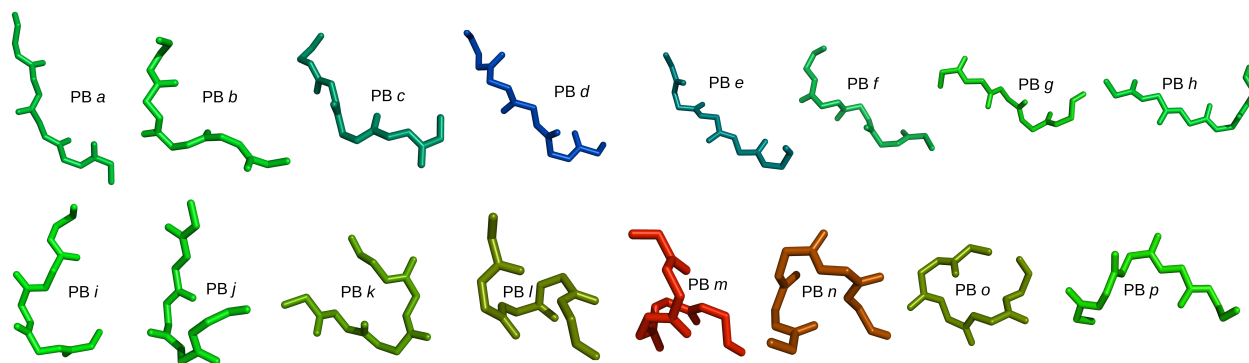
Proteins mediate all biological functions. The tight correlation between protein function and dynamics has led to molecular dynamics (MD) simulations becoming an indispensable tool in the bioinformatician's toolbox. A variety of molecular dynamics simulation softwares have been developed to this end. Although performing molecular dynamics (MD) of biological macromolecules like proteins is very computationally intensive, recent developments in computer hardware has enabled longer (micro second scale) simulations for larger systems [FRE\*06, OMO\*14, SGB\*14]. This results in the requirement of efficient and innovative tools for analyzing and visualizing MD trajectories.

Monitoring the protein backbone conformation i.e. orientation of its constituent atoms in 3D space, can be of interest for multiple reasons like to study domain movement, to study protein-ligand and protein-protein interfaces, to characterize disordered regions etc. Traditionally this is achieved by either visualizing 3D snapshots from the MD trajectory at regular time intervals or by assigning three state ( $\alpha$ -helix,  $\beta$ -strand & coil) or eight state ( $\alpha$ -helix,  $\beta$ -bridge,  $\beta$ -strand, 3-10 helix,  $\pi$ -helix, turn, bend & coil) secondary structure [HDS96, RC13]. Visualizing snapshots in 3D can be computationally demanding and impractical in many cases as in the case of long MD trajectories. Assigning secondary structure is fast and computationally efficient but can hide useful structural infor-

mation, especially from loop regions. Indeed, studies have shown the existence of "order in disorder" in loop regions by using structural alphabets [RMNC10].

Structural alphabets are libraries of recurrent structural patterns in proteins that can be used to approximate protein backbone representation. These libraries vary in both the size and number of the structural patterns. We use a structural alphabet called Protein Blocks (PBs) developed by de Brevern et al. [BEH00] due to its optimal balance of library size and the accuracy in approximating protein structures. PBs are a 16 state classification of the protein backbone based on dihedral ( $\phi$  &  $\psi$ ) angles. Each of these 16 structural patterns are 5 consecutive amino acids defined by eight consecutive  $\phi$ - $\psi$  values (Fig. 1). The 16 PBs are denoted by the alphabets  $a$  to  $p$ .

It is to be noted that PBs are not a sub-classification of secondary structure. For example, although PBs  $l, m, n$  occur very frequently in helices, they are not exclusive. That is to say occasionally PBs like  $k, o, .$  can occur in helices, hence capturing sub-secondary structure information like kinks and bends in helices. The versatility of PBs to abstract protein structures has been used in various examples [JAM\*10]. Thus, transforming individual frames from a MD simulation to PB sequences is an efficient way to represent 3D information in 2D. Here, we leverage on this representation to provide an efficient and scalable MD trajectory analysis tool.



**Figure 1:** Representative structures of the 16 structural patterns that comprise the structural alphabet Protein Blocks (PB) [BEH00]. PBs are colored based on the propensity coloring scheme.

## 2. Current state of the art

Recently, tools have been developed for both specific MD trajectory analysis e.g. analyzing protein-ligand interaction [DHR\*19], identifying cavities and channels in proteins [JBB\*18] and also for flexible and customisable analysis e.g. DIVE [RBD14] and VIA-MD [SLK\*18]. For an extensive review of specialized tools that have been developed to analyse MD trajectories, see section 6 of the article by Kozlikova et al. [KKF\*17]. Currently two tools are available that help in structural alphabet based analysis of MD trajectories- GSATools [PFFK13] and PBxplore [BSC\*17]. GSATools is based on a 25 state structural alphabet developed by Pandini et al. [PFK10]. It provides the user the ability to analyse correlated local motion and allostery from GROMACS MD trajectories. The structural alphabet used in GSATools was obtained by hierarchical clustering of the three angles describing the conformation of four consecutive backbone C $\alpha$  atoms [PFK10]. PBxplore is based on PBs and provides a useful representation of the different local conformations exhibited by each residue of the protein. It also provides an entropic measure called equivalent number of PBs ( $N_{eq}$ ) for identifying rigid and flexible regions in proteins [BSC\*17].

## 3. Overview and implementation

PBmapclust is a new PB based tool that represents changes in the protein backbone conformation during the MD simulation as PBmapclust plots and PB logos. Both these representations are complementary to each other. PBmapclust is computationally efficient and scalable, generating both the plot and the logos for a 200 frame MD trajectory in 20 seconds and for 2000 frames in 200 seconds on a quad-core processor with a clock speed of 3.30 GHz. Additionally, PBmapclust includes a novel clustering procedure to help in the identification of better representative structures from the MD simulation. Each of these features of PBmapclust are illustrated by considering the protein sba1, a HSP90 co-chaperone as a case study. The novelty of our method lies in encoding useful information in the coloring scheme of both the PBmapclust plot and the PB logos and also implementing PB substitution based clustering.

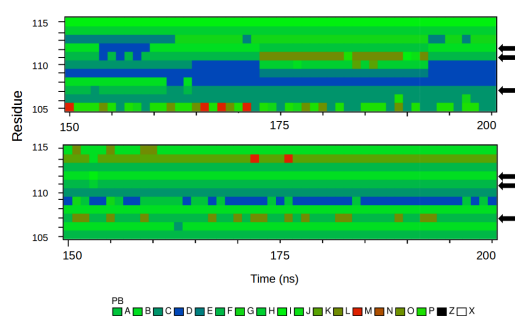
### 3.1. PBmapclust plot

PBmapclust generates a plot of MD frames vs. protein residues, color coded to match each of the 16 PBs. This plot is generated by a combination of Perl and Bash scripts that are part of the PBmapclust program. The program requires the GROMACS (GROningen MAchine for Chemical Simulations) [AMS\*15] run input file (.tpr) and trajectory file (.trr or .xtc) as inputs. It also requires the GROMACS and DSSP (Define Secondary Structure of Proteins) [KS83, JTK\*11] programs to be installed, both of which are free and open source. Although the GROMACS program is necessary for performing PBmapclust, trajectories from other molecular dynamics softwares can also be analysed using PBmapclust by converting them to either .xtc or .trr formats using the *mdconvert* tool from the MDTraj package [MBH\*15]. Currently PBmapclust runs only on the Linux platform. Generating a PBmapclust plot involves three main steps, namely:

1. Extracting all frames from the trajectory file.
2. Calculating torsion angles and PB assignment.
3. Generating a X PixelMap (XPM) file and subsequently, an Encapsulated PostScript (EPS) file.

A detailed explanation of each of these steps is included in the supplementary materials.

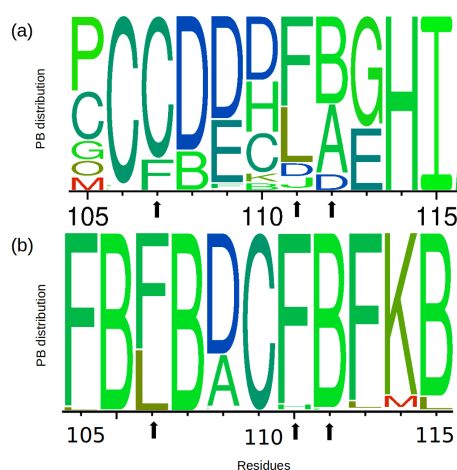
There are two color coding schemes implemented in PBmapclust. The *random* coloring scheme serves the sole purpose of being able to differentiate between structural changes, hence assigns random contrasting colors to consecutive PBs. The second scheme, *propensity* colors the PBs based on their three state secondary structure propensity. The proportion in which red:green:blue are mixed to arrive at the color for a particular PB is based on the  $\alpha$ -helix:coil: $\beta$ -strand propensity of that PB as described in de Brevern et al. [BEH00] and illustrated in Fig. 2. This helps us to identify drastic structural changes as more contrasting color changes. The user can change the coloring scheme by changing the parameters to the script. Thus the coloring scheme serves two purposes: representing the secondary structure tendencies of the PBs and also to highlight the underlying structural continuity that the PBs categorize.



**Figure 2:** Comparative analysis of the active site residue region for *sba1*-full (top) and *sba1*-core (bottom) during the last 50 ns of a 200 ns MD simulation. *Sba1* is a HSP90 co-chaperone protein, described in detail in section 4. The active site residues are indicated by arrows.

### 3.2. PB logos

PBmapclust plot represents the quantitative aspect of structural flexibility during the MD simulation but the qualitative aspect i.e. which PBs are changing at a particular position can be difficult to discern from this representation. To help with this analysis, as a complement to the PBmapclust plot, we provide a logo representation of the position wise PB occurrence (Fig. 3). The weblogo



**Figure 3:** PB logos of the active site residues of *sba1* (a) with the disordered domain (*sba1*-full) (b) without the disordered domain (*sba1*-core).

program [CHCB04] is used to generate the PB logos. Weblogo is a tool originally developed to visualize multiple sequence alignments. Here we have adopted weblogo for PB variations during the MD simulation. The height of the PB corresponds to the frequency of occurrence of the PB. For example if PB *m* occurs in 70 frames out of the 100 frames of the MD simulation and PB *n* occurs in the remaining 30 frames, then the height of the logo is 70% PB *m* and 30% PB *n*. The number of different PBs occurring at a particular position is proportional to the structural flexibility of that position. This representation is similar to the logos generated by PBxplore

but here the user has the option to choose the coloring scheme—*random* or *propensity*.

### 3.3. Clustering of conformations

It is a common practice to perform clustering and identify representative structures from a MD simulation to expedite further analysis e.g. molecular docking. In this section, we describe a novel PB substitution based clustering method.

In 1983, Kimura described a metric to study molecular evolution in DNA by differentiating the purine and pyrimidine substitutions [Kim83]. We have adopted a similar metric to differentiate between favorable and unfavorable PB substitutions based on a previously established PB substitution matrix [TGS\*06]. All possible pairwise distances between conformations extracted from the MD trajectory were calculated. For every pair of PB sequence each PB pairs were scored based on the substitution matrix. Positions with positive scores (*m*) are favored transitions while those with negative scores are unfavored transitions. Negatively scored regions basically indicate structurally variable regions. Between every pair of PB sequences, the Kimura distance [Kim83] was calculated as follows :

$$\text{Kimura distance} = -\ln(1 - D - 0.2D^2)$$

where,

$$D = 1 - S \quad \text{and} \quad S = \frac{m}{n_{pos}}$$

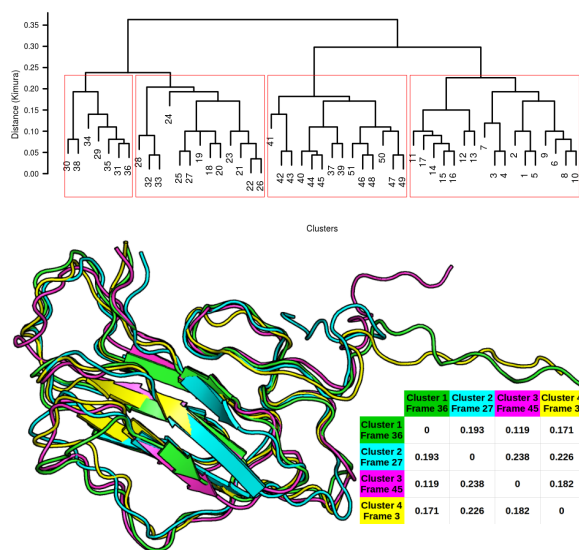
*m* - number of positions with positive scores  
*n<sub>pos</sub>* - number of positions scored

A symmetrical dissimilarity matrix was populated with these distances. This was further analysed using the *hclust* algorithm implemented in R [RC18]. The four clusters obtained by the above procedure and their representative structures for the case study protein-*sba1* are shown in Fig.4.

### 4. Case study: *sba1*- A HSP90 co-chaperone

Heat shock proteins (HSP) are a family of proteins ubiquitously found across all living organisms to help them tide over stress conditions, HSP90 being the most widely studied of them all. *Sba1* is one of the co-chaperones to the HSP90 chaperoning machinery in *Candida albicans*, a fungal pathogen. Inhibiting the chaperone-co-chaperone interaction has been proposed as an efficient strategy to curb fungal infection [CFW\*10].

Due to the absence of an available crystal structure for the *C. albicans sba1*, we rely on homology modeling. The N-terminal domain can be modeled reliably due to homologous structures from *Saccharomyces cerevisiae* (PDB ID 2CG9) [AMV\*06] and human (PDB ID 1EJF) [WSF\*00] co-chaperones. But the C-terminal domain poses a problem due to being majorly disordered [WAB99]. More importantly, it has been shown that the C-terminal plays an important role in the binding of gedunin, an inhibitor of chaperone-co-chaperone interaction [PPF\*13]. Here we illustrate the utility of PBmapclust in highlighting the structural differences between the homology models with and without the disordered region. The



**Figure 4:** Top: Clustering of 50 conformations from the last 50 ns of a 200 ns MD simulation of *sbal-core*. The clustering is based on Kimura distances. Bottom: Superposition of the representative structures with kimura distances in insert.

model for the *sbal-core* (without C-terminal disordered domain) was built using the *S. cerevisiae* and human co-chaperones as templates using the Prime program from Schrödinger LLC [Sch18]. The model for the entire protein, with the disordered domain was built using the Phyre2 [KMY\*15] homology modelling webserver. Based on residue conservation when compared to the human co-chaperone, Thr107, Leu111 and Lys112 have been hypothesized to be the active site residues, hence we will focus on this region of the protein.

The *sbal-core* and *sbal-full* models were subjected to 200 ns MD simulation in explicit TIP3P water solvent using the CHARMM36m [HRN\*17] forcefield in GROMACS. From the RMSD (Root Mean Squared Deviation) plot in the supplementary materials, we observe that both structures exhibit significant structural variations during the MD simulation but the *sbal-core* stabilizes after 150 ns at approximately 10 Å RMSD from its starting conformation. The *sbal-full* structure still shows an increasing RMSD trend, but such a feature was expected due to the presence of the large disordered region. Hereafter, we will focus on the last 50 ns of the MD simulation.

The PBmapclust plots in Fig. 2 are for residues 105 to 115, which covers the residues of interest i.e. Thr107, Leu111 and Lys112 for the last 50 ns of the simulation. The plots and PB logos for the entire protein are included in the supplementary materials. In both the plots, we see a predominance of PBs corresponding to the loop and coil regions, indicated by the green shades. In the case of *sbal-full*, there is slightly higher tendency towards  $\beta$ -strand, indicated in blue. Globally, there is significantly more structural variation in the active site of *sbal-full* than compared to the *sbal-core*. PBs *c* & *f* instead of *f* & *l* for Thr107, PBs *f* & *l* interspersed with PBs *d* & *j* vs PB *f* for Leu111 and PBs *b*, *a* & *d* vs PB *b* for Lys112.

Thus the disordered region does actually affect the conformation of active site residues. Such information can be used to develop a rational method for selecting protein conformations from the MD trajectory for further analysis, like in this case to perform molecular docking.

## 5. Interpretation of output

PBmapclust was developed with the intention of being a complementary tool to classical MD trajectory analysis like RMSD plot, RMSF (Root Mean Squared Fluctuation) plot, secondary structure analysis etc. While RMSD and RMSF are global metrics averaged over the entire protein and the entire duration of the MD respectively, PBmapclust allows the user to go into the specifics. For example, after a specific duration of MD, the RMSD could be stable when looking at the entire protein but still have significant deviation for a loop region of interest. Such features can be easily identified as color changes on the PBmapclust plot. Hence, the interpretation of the PB logos will be more relevant if it is generated from a time step after which the system is stabilized, which can be determined from the RMSD plot. Although individual scenarios vary, in general it might be useful to use the *propensity* coloring scheme to get a global overview of the entire MD trajectory before using the *random* scheme. This way, the regions of explicit structural variation are easily identified and then the shades of red, blue and green can be further highlighted by the contrasting colors of the *random* coloring scheme.

## 6. Conclusions

PBmapclust is a simple tool that can be used synergistically not only with the classical MD analysis tools like RMSD and RMSF plots but also with PBxplore. The clustering feature enables the user to choose conformations from the MD simulation that are more representative of the structure space and avoid redundancy. Further, representing MD trajectories as images opens up the avenue for the use of powerful image comparison tools for MD trajectory comparison. This can be useful for comparing multiple MD runs for the same protein like *apo* and *holo* forms of the protein, comparing replicate runs etc.

All the scripts to perform PBmapclust are available at <https://github.com/iyandarvetrivel/PBmapclust> under the GNU general public license.

## Acknowledgements

Part of this work was supported by a grant to BO in the framework of the GRIOTE project funded by Région Pays de la Loire (France). IV and ADL thank the INCa financial support of the PLBIO grant. SG and ADL acknowledge the Région Pays de la Loire (Dynamique scientifique PIRAMID) for the support. This research used the local computer resources from the CCIPL (Centre de Calcul Intensif des Pays de Loire).

## References

[AMS\*15] ABRAHAM M. J., MURTOLA T., SCHULZ R., PÁLL S., SMITH J. C., HESS B., LINDAHL E.: GROMACS: High performance

- molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1 (2015), 19–25. 2
- [AMV\*06] ALI M. M., MARK ROE S., VAUGHAN C. K., MEYER P., PANARETOU B., PIPER P. W., PRODRIMOU C., PEARL L. H.: Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* (2006). 3
- [BEH00] BREVERN A. G. D., ETCHBEST C., HAZOUT S.: Bayesian Probabilistic Approach for Predicting Backbone. *Proteins* 287, January (2000), 271–287. 1, 2
- [BSC\*17] BARNOUD J., SANTUZ H., CRAVEUR P., JOSEPH A. P., JALLU V., DE BREVERN A. G., POULAIN P.: PBxplorer: a tool to analyze local protein structure and deformability with Protein Blocks. *PeerJ* 5 (nov 2017), e4013. 2
- [CFW\*10] CHADLI A., FELTS S. J., WANG Q., SULLIVAN W. P., BOTUYAN M. V., FAUQ A., RAMIREZ-ALVARADO M., MER G.: Celestrol Inhibits Hsp90 Chaperoning of Steroid Receptors by Inducing Fibrillization of the Co-chaperone p23. *Journal of Biological Chemistry* 285, 6 (feb 2010), 4224–4231. 3
- [CHCB04] CROOKS G. E., HON G., CHANDONIA J. M., BRENNER S. E.: WebLogo: A sequence logo generator. *Genome Research* (2004). 3
- [DHR\*19] DURAN D., HERMOSILLA P., ROPINSKI T., KOZLIKOVA B., VINACUA A., VAZQUEZ P.-P.: Visualization of Large Molecular Trajectories. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (jan 2019), 987–996. 2
- [FRE\*06] FITCH B. G., RAYSHUBSKIY A., ELEFThERIOU M., WARD T. J. C., GIAMPAPA M., ZHESTKOV Y., PITMAN M. C., SUITS F., GROSSFIELD A., PITERA J., SWOPE W., ZHOU R., FELLER S., GERMAIN R. S.: Blue Matter: Strong Scaling of Molecular Dynamics on Blue Gene/L. In *Computational Science – ICCS 2006* (Berlin, Heidelberg, 2006), Alexandrov V. N., van Albada G. D., Sloot P. M. A., Dongarra J., (Eds.), Springer Berlin Heidelberg, pp. 846–854. 1
- [HDS96] HUMPHREY W., DALKE A., SCHULTEN K.: VMD: Visual molecular dynamics. *Journal of Molecular Graphics* (1996). 1
- [HRN\*17] HUANG J., RAUSCHER S., NAWROCKI G., RAN T., FEIG M., DE GROOT B. L., GRUBMÜLLER H., MACKERELL A. D.: CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14, 1 (jan 2017), 71–73. 4
- [JAM\*10] JOSEPH A. P., AGARWAL G., MAHAJAN S., GELLY J.-C., SWAPNA L. S., OFFMANN B., CADET F., BORNOT A., TYAGI M., VALADIÉ H., SCHNEIDER B., ETCHBEST C., SRINIVASAN N., DE BREVERN A. G.: A short survey on protein blocks. *Biophysical reviews* 2, 3 (aug 2010), 137–147. 1
- [JBB\*18] JURCIK A., BEDNAR D., BYSKA J., MARQUES S. M., FURMANOVA K., DANIEL L., KOKKONEN P., BREZOVSKY J., STRNAD O., STOURAC J., PAVELKA A., MANAK M., DAMBORSKY J., KOZLIKOVA B.: CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* 34, 20 (oct 2018), 3586–3588. 2
- [JTK\*11] JOOSTEN R. P., TE BEEK T. A., KRIEGER E., HEKKELMAN M. L., HOOFT R. W., SCHNEIDER R., SANDER C., VRIEND G.: A series of PDB related databases for everyday needs. *Nucleic Acids Research* (2011). 2
- [Kim83] KIMURA M.: *The neutral theory of molecular evolution*. Cambridge University Press, 1983. 3
- [KKF\*17] KOZLÍKOVÁ B., KRONE M., FALK M., LINDOW N., BAADEN M., BAUM D., VIOLA I., PARULEK J., HEGE H. C.: Visualization of Biomolecular Structures: State of the Art Revisited. *Computer Graphics Forum* (2017). 2
- [KMY\*15] KELLEY L. A., MEZULIS S., YATES C. M., WASS M. N., STERNBERG M. J.: The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* (2015). 4
- [KS83] KABSCH W., SANDER C.: Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 12 (1983), 2577–2637. 2
- [MBH\*15] MCGIBBON R. T., BEAUCHAMP K. A., HARRIGAN M. P., KLEIN C., SWAILS J. M., HERNÁNDEZ C. X., SCHWANTES C. R., WANG L. P., LANE T. J., PANDE V. S.: MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* (2015). 2
- [OMO\*14] OHMURA I., MORIMOTO G., OHNO Y., HASEGAWA A., TAJI M.: MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372, 2021 (jun 2014), 20130387–20130387. 1
- [PFFK13] PANDINI A., FORNILI A., FRATERALI F., KLEINJUNG J.: GSATools: analysis of allosteric communication and functional local motions using a structural alphabet. *Bioinformatics* 29, 16 (aug 2013), 2053–2055. 2
- [PFK10] PANDINI A., FORNILI A., KLEINJUNG J.: Structural alphabets derived from attractors in conformational space. *BMC Bioinformatics* 11, 1 (feb 2010), 97. 2
- [PPF\*13] PATWARDHAN C. A., FAUQ A., PETERSON L. B., MILLER C., BLAGB B. S. J., CHADLI A.: Gedunin inactivates the co-chaperone p23 protein causing cancer cell death by apoptosis. *The Journal of biological chemistry* 288, 10 (mar 2013), 7313–25. 3
- [RC18] R CORE TEAM: R: A Language and Environment for Statistical Computing, 2018. 3
- [RBD14] RYSAVY S. J., BROMLEY D., DAGGETT V.: DIVE: a graph-based visual-analytics framework for big data. *IEEE computer graphics and applications* 34, 2 (2014), 26–37. 2
- [RC13] ROE D. R., CHEATHAM T. E.: PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* (2013). 1
- [RMNC10] REGAD L., MARTIN J., NUEL G., CAMPROUX A. C.: Mining protein loops using a structural alphabet and statistical exceptionality. *BMC Bioinformatics* (2010). 1
- [Sch18] SCHRÖDINGER LLC: Prime, version 2018-2, 2018. 4
- [SGB\*14] SHAW D. E., GROSSMAN J., BANK J. A., BATSON B., BUTTS J. A., CHAO J. C., DENEROFF M. M., DROR R. O., EVEN A., FENTON C. H., FORTE A., GAGLIARDO J., GILL G., GRESKAMP B., HO C. R., IERARDI D. J., ISEROVICH L., KUSKIN J. S., LARSON R. H., LAYMAN T., LEE L.-S., LERER A. K., LI C., KILLEBREW D., MACKENZIE K. M., MOK S. Y.-H., MORAES M. A., MUELLER R., NOCILOLO L. J., PETICOLAS J. L., QUAN T., RAMOT D., SALMON J. K., SCARPAZZA D. P., SCHAFER U. B., SIDDIQUE N., SNYDER C. W., SPENGLER J., TANG P. T. P., THEOBALD M., TOMA H., TOWLES B., VITALE B., WANG S. C., YOUNG C.: Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis* (nov 2014), IEEE, pp. 41–53. 1
- [SLK\*18] SKÁNBERG R., LINARES M., KÖNIG C., NORMAN P., JÖNSSON D., HOTZ I., YNNERMAN A.: VIA-MD: Visual Interactive Analysis of Molecular Dynamics. In *Workshop on Molecular Graphics and Visual Analysis of Molecular Data* (2018), Byska J., Krone M., Sommer B., (Eds.), The Eurographics Association. 2
- [TGS\*06] TYAGI M., GOWRI V. S., SRINIVASAN N., DE BREVERN A. G., OFFMANN B.: A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins: Structure, Function and Genetics* (2006). 3
- [WAB99] WEIKL T., ABELMANN K., BUCHNER J.: An unstructured C-terminal region of the Hsp90 co-chaperone p23 is important for its chaperone function. *Journal of Molecular Biology* 293, 3 (1999), 685–691. 3
- [WSF\*00] WEAVER A. J., SULLIVAN W. P., FELTS S. J., OWEN B. A., TOFT D. O.: Crystal structure and activity of human p23, a heat shock protein 90 co-chaperone. *The Journal of biological chemistry* 275, 30 (jul 2000), 23045–52. 3