A New Look at the Structures of Old Sepsis Actors by Exploratory Data Analysis Tools

Antonio Gnoni [1] , **Emanuele De Nitto** [1] , **Salvatore Scacco** [1] **and Luigi Santacroce** [2,*]
and Luigi Leonardo Palese [1,*]

[1] SMBNOS—Università degli Studi di Bari, 70124 Bari, Italy; antonio.gnoni@uniba.it (A.G.);
   emanuele.denitto@uniba.it (E.D.N.); salvatore.scacco@uniba.it (S.S.)

[2] Ionian Department, Microbiology and Virology Lab Unit, University Hospital of Bari, Università degli Studi
   di Bari, 70124 Bari, Italy

[*] Correspondence: luigi.santacroce@uniba.it (L.S.); luigileonardo.palese@uniba.it (L.L.P.)

Abstract: Sepsis is a life-threatening condition that accounts for numerous deaths worldwide, usually complications of common community infections (i.e., pneumonia, etc), or infections acquired during the hospital stay. Sepsis and septic shock, its most severe evolution, involve the whole organism, recruiting and producing a lot of molecules, mostly proteins. Proteins are dynamic entities, and a large number of techniques and studies have been devoted to elucidating the relationship between the conformations adopted by proteins and what is their function. Although molecular dynamics has a key role in understanding these relationships, the number of protein structures available in the databases is so high that it is currently possible to build data sets obtained from experimentally determined structures. Techniques for dimensionality reduction and clustering can be applied in exploratory data analysis in order to obtain information on the function of these molecules, and this may be very useful in immunology to better understand the structure-activity relationship of the numerous proteins involved in host defense, moreover in septic patients. The large number of degrees of freedom that characterize the biomolecules requires special techniques which are able to analyze this kind of data sets (with a small number of entries respect to the number of degrees of freedom). In this work we analyzed the ability of two different types of algorithms to provide information on the structures present in three data sets built using the experimental structures of allosteric proteins involved in sepsis. The results obtained by means of a principal component analysis algorithm and those obtained by a random projection algorithm are largely comparable, proving the effectiveness of random projection methods in structural bioinformatics. The usefulness of random projection in exploratory data analysis is discussed, including validation of the obtained clusters. We have chosen these proteins because of their involvement in sepsis and septic shock, aimed to highlight the potentiality of bioinformatics to point out new diagnostic and prognostic tools for the patients.

Keywords: Sepsis; allosteric; albumin; cyclooxygenase; hemoglobin; PCA; random projection; pathophisiology; bioinformatics tools; clinical chemistry

---

1. Introduction

Sepsis is a life-threatening condition that accounts for numerous deaths worldwide, usually as complications of community infections (i.e., pneumonia, etc.), or infections in hospitalized patients. Sepsis and septic shock, its most severe evolution, involve the whole organism, recruiting and producing a lot of molecules, mostly proteins. Protein functions are closely related with their structure, and the discovery of meaningful structure-function relationships is of overwhelming importance in biochemistry. Conformational changes in proteins have been known for a long time and are crucial for the biological activity of these molecules. These changes range from subtle side-chain displacement

or change in the flexibility in some loop to large whole domain motions. Conformational changes have been involved in the enzymatic activities of proteins, in the recognition of substrates and in the protein-protein interactions. Because of their importance, numerous experimental and computational techniques were developed to allow the extensive characterization of these conformational changes so that it is virtually impossible to remember them all here. In recent years there has been a considerable increase in the ability to produce high-quality three-dimensional structures of proteins. To date more than 126,000 structure are in the Protein Data Bank (PDB) [1–3]. This number continues to grow dramatically and for many proteins multiple entries are present in the PDB. Important information about conformational states of specific proteins can be extracted by the analysis of these redundant entries for the same protein, generally obtained in different conditions. For single protein, redundant data sets can be analysed using various mathematical tool [4]. A classical approach is the principal component analysis (PCA), a multivariate statistical method based on the covariance of data [5–7]. This method has a wide range of applications in today's data science [8–11]. If the number of data (or different structures in the case of proteins) is sufficiently high, PCA even makes it possible to reconstruct the main modes of protein motion starting from the (static) crystallographic structures, in excellent agreement with the experimental and molecular dynamics data [12,13]. However, the fact that the number of entries in these structural data sets is large but in general not comparable (i.e., less than) the number of degrees of freedom that are needed to describe a typical protein imposes several constraint to the algorithm to be used in such analyses. In this case, to perform a PCA type analysis, it is necessary to use specialized state of the art algorithms [14], which are also able in this type of data sets to reveal information on the dynamics [15] or the presence of functionally important clusters [16].

The reader should consider that crystal structures represent time and space averages of all molecules present within the crystal lattice (which is not perfect). Conformational variations can provide information about the flexibility or movement of regions of protein structure that might be important for function and ligand binding. Even in the case of a single structure corresponding to the minimum of a potential well, the protein is actually a family of structures that can be explored as a result of thermal motion. Particularly in the case of subtle structural differences it is necessary to consider not only if and how important are these, or are they related to some functional aspect of the protein, but first of all if they can be simply due to thermal motion (so to speak, frozen in the coordinates provided in the PDB), or also to refinement errors [17].

Here we show that a simple algorithm based on random projection [18] performs well in the dimensionality reduction and unsupervised clustering of protein structure data sets. Furthermore, if data clusters are effectively well separated, this will be true even in the case of random projection. Therefore, if we find clusters of data in two-dimensional projections obtained by PCA that are not observable even in the random projection, it is possible that the clusters are not reliable. In this case caution is required in the interpretation of the data, which must be integrated with the biochemical knowledge available on the particular proteins. We apply this algorithm in the exploratory data analysis on three model proteins that represent different types of allostery from a structural point of view: a monomeric allosteric protein that exhibits evident structural changes, a case of allostery without dramatic structural changes, and a classical multimeric allosteric protein. All these proteins are involved in various ways in sepsis and, to better understand this process, the study of the conformational changes of existing and newly produced proteins that occur during an infectious process is really interesting.

2. Results

2.1. The Human Serum Albumin: Allostery in a Monomer

The human serum albumin (HSA) [19], the most abundant protein in plasma, is a monomeric multi-domain molecule. HSA is a non-glycosylated, all-α protein chain of 65 kDa, with a globular heart-shaped conformation containing three homologous domains. Each domain is composed by two

subdomains. It is an important transport protein with different binding sites able to accommodate a number of chemically different ligands. HSA represents the main carrier for fatty acids, for which there are seven binding sites. It is also a depot and carrier for exogenous compounds (mainly, but not exclusively at the so called Sudlow's sites I and II), thus affecting the pharmacokinetics of many drugs. Hypoalbuminaemia is often associated with sepsis and/or critical illness, and the supplementation of HSA still remains controversial in these patients [20]. In fact, the function of HSA is fundamental in the infective and septic process, and is closely related to specific conformational modifications, influencing the whole health status of the patients [21]. It is worth noting that a large number of structural and functional works on HSA have lead to the conclusion that two structures, possibly related to the presence of fatty acids, are discernible for this protein [19,22]. Short chain fatty acids (SCFAs), a common product of microbial metabolism, affect albumin production and metabolism, so they have a role in the evolution of the septic patients [23]. In fact, they directly influence the hepatic albumin metabolism [24]. This three-domain organization of HSA is at the root not only of its extraordinary ligand binding capacity, but also of the allosteric control of this last. The HSA structure and reactivity (and also its enzymatic activity) is affected reversibly by pH and ligands, such as fatty acids, heme or drugs.

Among the available structures, we selected 58 structure for the analysis. This data set has been described in details elsewhere [18]. The α-carbon atom Cartesian coordinates of HSA were extracted and arranged in a data matrix, such that each row represented a single HSA structure. Thus, the data matrix was composed of 58 rows and 1695 columns (565 α-carbon atoms were finally included in the analysis [18]). This is a degenerated data set, as it is impossible to obtain the true correlation matrix of a multivariate system with 1695 degree of freedom by using only 58 samples. As recalled above, in order to reduce the dimensionality and to obtain an unsupervised clustering of the structures present in the data set, it is possible to use algorithms that estimate the principal components. Using the truncated singular value decomposition (SVD) algorithm [14] to estimate the principal components, two clusters of structures for the HSA data set can be discerned, as can be seen from Figure 1. However, the same clusters can be obtained by the simple random projection algorithm. As can be easily appreciated by inspecting the figure, these analyses clearly demonstrate that the only discriminant for such a structural switch in the whole data set is the presence or absence of bound fatty acids [18].
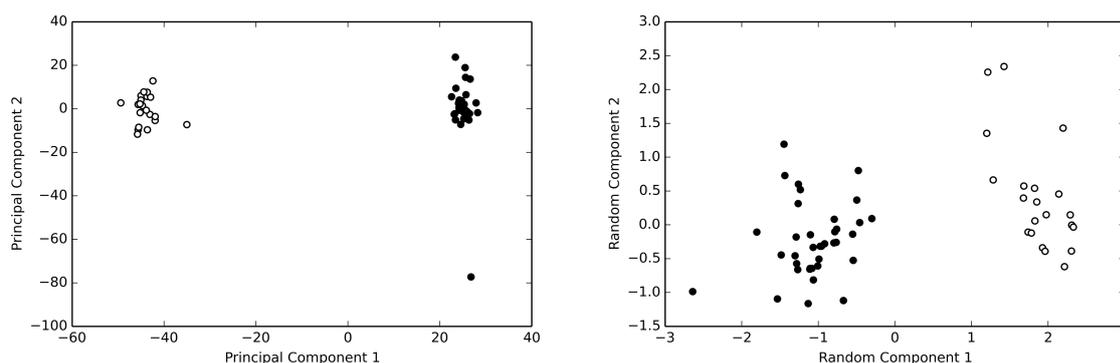

Figure 1. The human serum albumin data set. Principal component analysis (**left panel**) and random projection analysis (**right panel**) of the of the normalized HSA data set are reported. HSA structures without bound fatty acids are reported as withe circles, while those with bound fatty acids are reported as black circles. Both methods clearly allow to recognize two clusters of structures. Note the different level of dispersion provided by the two methods.

2.2. The Cyclooxygenase: Allostery without Conformational Change

The cyclooxygenase (COX), also known as prostaglandin H_2 (PGH$_2$) synthase or prostaglandin endoperoxide H_2 synthase (PGHS), is a membrane bound, heme-dependent bis-oxygenase and hydroperoxidase [25–27]. This enzyme participates to the prostanoid synthesis by two sequential

reactions: the bis-oxygenation of arachidonic acid (the cycloxigenase reaction) and the reduction of prostaglandin G2 (PGG$_2$) (in the peroxidase site) to form PGH$_2$. In mammals, arachidonic acid is the major prostanoid precursor, which are a subclass of the eicosanoids. COX has a pivotal role in the production of a large number of immune and inflammatory mediators, and the effectiveness of COX inhibition as a treatment for severe sepsis has been extensively studied [28]. Two isoforms of COX can be found in mammals, the constitutive COX-1 and the inducible COX-2. These two isoforms are significantly different in their expression profiles and physiological roles and are involved in various pathological situations. From a structural point of view, and as expected considering the sequence similarity, the two isoforms are quite similar. COX functions as homodimer, and each monomer consists of three domains [26]: an EGF domain at the N-terminal, a membrane-binding domain and a large globular C-terminal domain. This last domain contains the heme binding site and is the responsible of the catalytic activities of these enzymes. The EGF domain participates to the dimer interface and probably to the interaction with membranes. The membrane-binding domain consists of four short amphipathic α-helices. The bulk of COX is represented by the catalytic domain, which is composed essentially by α-helices. Nonsteroidal anti-inflammatory drugs (NSAIDs) are a drug class that inhibit the COX activity. NSAIDs can be divided in two classes: the classical isoform non-specific, that inhibit both COX-1 and COX-2, and the COX-2 inhibitors show a high selectivity for this particular isoform. A large number of studies has demonstrated that COX is a dynamic and flexible molecule that does undergo conformational changes upon binding of heme, substrates and drugs [26].

The fact that COX works as a homodimer and a series of data on its enzymatic activity strongly suggest that this enzyme can undergo to allosteric regulation by its substrates [29–31]. However, despite a growing number of crystal structures available in different conditions, no evident ligand-induced conformational changes can be noticed [26]. We have analysed data sets of these enzymes as example of proteins where only tiny (if any) structural changes can be observed. We selected 38 *Ovis ares* COX-1 structures and from these we obtained a 38 × 1653 matrix representing the Cartesian coordinates of the α-carbon atoms (551 α-carbon atoms). The COX-2 data set included 78 entries from the *Mus musculus* specie, arranged in a matrix of dimension 78 × 1608 (536 α-carbon atoms).

The results of the PCA (by the truncated SVD methods recalled above) and the random projection analysis for the COX-2 data set are reported in Figure 2. Both methods show that all the analyzed structures are distributed in a single cluster, in agreement with what is known about the structural variability of this enzyme in different conditions. PCA analysis detects some putative outliers, indicated as gray circles in the Figure, which are located at the peripheral region of the cluster obtained by the random projection algorithm, but not linearly separable from the bulk of structures. Moreover, the outliers distribution is not exactly the same using the two methods: this suggests that, in this case, the separation obtained by PCA is probably strongly influenced by the noise due to the low number of available samples. It should be noted that no meaningful partition of these data can be obtained considering the presence (or absence) of ligands, such as NSAIDs, fatty acids or heme, in agreement with the conclusion that probably only one cluster of structures is actually present in the data set.

The *Ovis ares* COX-1 data set shows different results, depending on the algorithm used for the dimensionality reduction. As indicated by Figure 3, the PCA algorithm describes three different clusters of structures (labeled as red, black and green circles in the Figure). One of this, the one shown in green in the Figure 3, is particularly interesting because it is composed by entries that have been crystallized as monomers with bound fatty acids [32–35]. The other two clusters are both composed by unliganded molecules or structures containing bound NSAIDs.
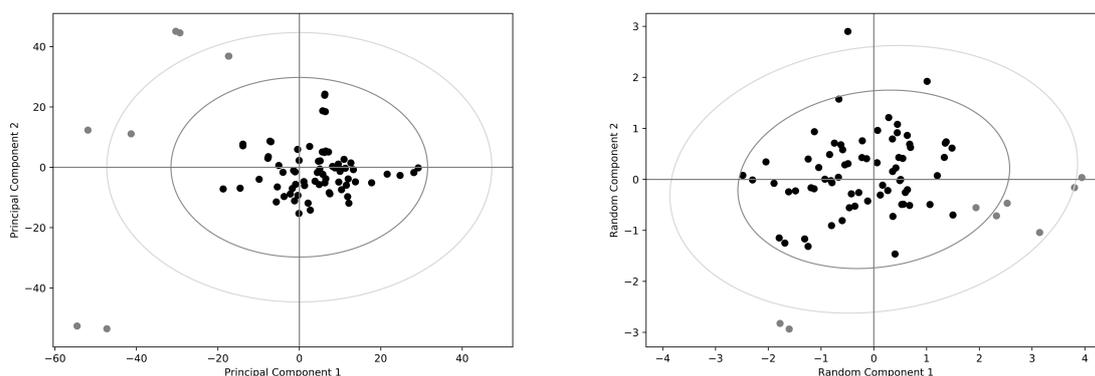
Figure 2. The *Mus musculus* COX-2 data set. Principal component analysis (**left panel**) and random projection analysis (**right panel**) are reported of the normalized data set containing the COX-2 monomers. Outliers in the principal component analysis are reported as gray circles in both panels; 95% and 99% confidence levels are drawn in dark and light gray, respectively. Note that both methods do not identify clearly separate clusters of structures.
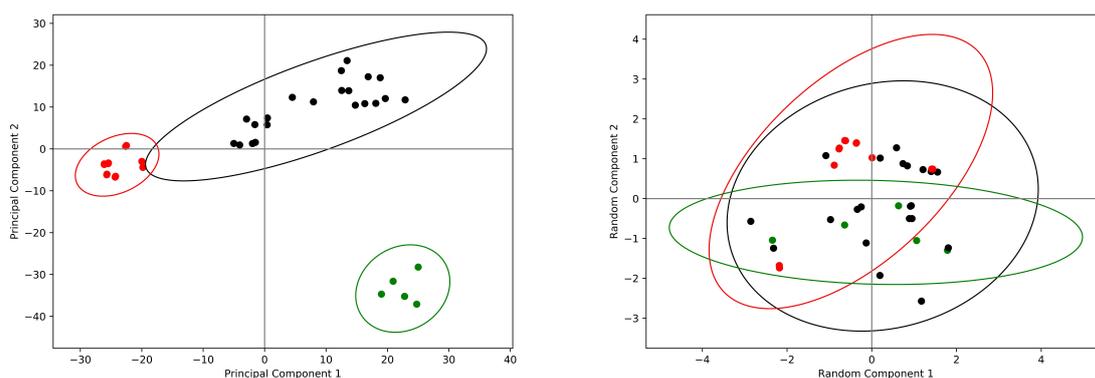

Figure 3. The *Ovis aries* COX-1 data set. Principal component analysis (**left panel**) and random projection analysis (**right panel**) are reported of the normalized data set containing the COX-1 monomers. Principal component analysis identifies on this data set three clusters of structures, indicated as black, red and withe circles. On the contrary the random projection method returns for this data set a single cluster; the entries are colored as in the left panel. The 99% confidence levels for the clusters are reported are drawn in the same color.

The random projection algorithm shows for this data set only a single cluster of structure, in which the structures that appeared in different cluster after PCA appear instead mixed. It should be noted that the differences between structures that are reported in different clusters by PCA are extremely small. In Figure 4 the superposition of the structures belonging to the cluster of structures with bound fatty acids appears almost perfect. In the same Figure 4 it is reported also a COX-1 structure that is very distant from those mentioned above [36]. As can be appreciated by inspecting the Figure, the differences are really minimal, so much so that it is not inconceivable that the PCA algorithm has operated a kind of over fitting of this data set. This observation is supported by the fact that the separation in different clusters vanishes in the two-dimensional random projection, suggesting also in this case that in reality only one cluster is present.
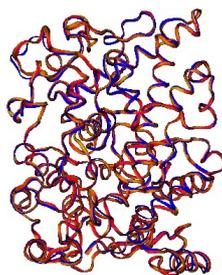
Figure 4. COX-1 structures. The figure reports the structures of PDB entries 1IGZ, 1IGX, 1FE2, 1U67 and 1DIY, that represent the cluster of structures with bound fatty acids described in the text (red and orange structures). The structure of the 1PGF (chain A) is reported in blue for comparison. All these structures belong to E.C 1.14.99.1, prostaglandin-endoperoxide synthase.

2.3. Hemoglobin: The Quintessence of Allostery

Hemoglobin (Hb) is undoubtedly the archetype of allosteric proteins [37,38]. Human adult hemoglobin (Hb A) has a tetrameric structure consisting of two α-chains and two β-chains with 141 and 146 amino acids respectively. Each of the chains in Hb contains a heme group, which is the binding site for ligands, such as oxygen, carbon monoxide, cyanide and nitric oxide. Hemoglobin usually drops in septic patients, due to a large number of factors, most of them still now undefined. The concentration of Hb in blood samples is currently accepted as a potent prognostic marker [39–43]. It is one of the first proteins whose structure was resolved by X-ray crystallography since the 1960s [44]. From these crystallographic data the Peruz's two-structure and the Monod-Wyman-Changeux models for Hb allostery were proposed [45–47]. These classical models essentially postulates that the four subunits in Hb assume simultaneously either the tense (T) or relaxed (R) structures. Both structures can bind ligands but the affinity towards the ligands changes at the transition from the T to the R form. The differences in the observed crystal structures of the Hb in its oxy- and deoxy- forms are correlated with the T- and R- states of the Monod-Wyman-Changeux model.

Hb can be considered a dimer of $\alpha\beta$ dimers. The two $\alpha\beta$ dimers are in contact and assume a two-fold symmetry with the symmetry axis passing trough a water filled cavity composed by the four subunits. The helices B, G and H (the BGH frame) form a well packed structure that does not change upon ligand binding. The C and G helices and the FG corner of the unlike subunits make the sliding contacts that change upon oxygen binding. The classical results of Perutz suggested that upon oxygenation the $\alpha_2\beta_2$ dimer rotates relatively to the other dimer, the heme Fe(II) moves through the porphyrin plane and several several inter-subunit and intra-subunit salt bridges are broken. Actually dozen of different structures of Hb are available and the clustering and classification of these is still an active research field. Obviously we are not interested here in a systematic analysis of all these structure, but simply to a comparison between different methods of unsupervised clustering. However it should be mentioned that these systematic analyses have shown that what emerges is significantly more complicated than the simple two state model for the Hb structure [48].

We have included in the analysis only 30 Hb tetramer. The selection criterion was simply based on the search for the structure with the highest rank in the in the PDB cluster containing the α-chain of the human Hb A, using a 100% identity cutoff, and the constraint of exactly a tetramer presents in the structure and the absence of multiple coordinates for the same α-carbon atom in the pdb file. The structures are represented by a 30×1722 matrix.

The results of these analyses are reported in Figure 5. As can be appreciated, the Hb structures form two distinct groups in the two-dimensional projections, both in that obtained by means of the truncated SVD algorithm and in that one obtained from the random projection algorithm.
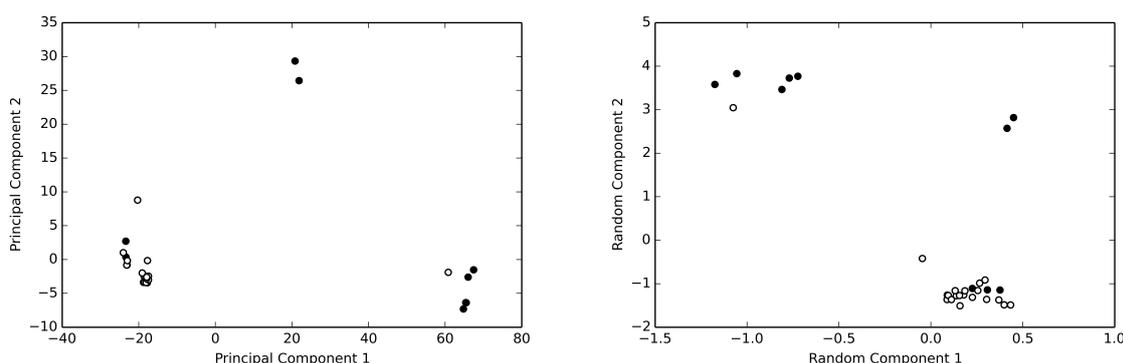
Figure 5. The human hemoglobin data set. Principal component analysis (**left panel**) and random projection analysis (**right panel**) are reported of the normalized data set containing the hemoglobin tetramers. Entries in this data set are reported as black circles if they represent liganded forms of the hemoglobin, or as withe circles if they are unliganded species. Two clusters of structures and two outliers are clearly detected by both algorithms. See text for further details.

These two clusters correspond essentially to liganded and unliganded forms of the Hb, with few, and perfectly explicable, exceptions. The structures 1QSI, 1THB and 1YE2 [49–51], although representing liganded forms of the Hb molecule, cluster with the deoxy-Hb if analysed by both algorithms. However all the remaining structures represent T state of the molecule. Moreover 1SHR [52] clusters with the liganded forms of the Hb, despite being a deoxy-Hb, using both algorithms. Its particularity is justified considering that it is the structure of the Hb A_2 with ferrocyanide bound. Interestingly, both algorithms report two structure as a separate mini-cluster, distinct both from the cluster containing the liganded structures and from that formed by the unliganded ones. These two structures (1SDK and 1SDL) have been obtained by using the trimesic acid for the avowed purpose of trapping the intermediates of the transition between the T form and the form R [53].

3. Discussion

In this work we have compared the effectiveness in dimensionality reduction for exploratory data analysis of two different algorithms. Both are capable to deal with degenerated data sets, i.e., data sets whose number of entries is much smaller than the number of the degrees of freedom that are required to describe the system. The first one is the truncated SVD method for the calculation of PCA [14], whereas the second one relies on random projection [18] which is based on the properties of random matrices [54–57] and the features of correlation matrices obtained from the protein dynamics [54–56,58,59].

The results of these analyses show that both algorithms are effective in the dimensionality reduction task, as well as the related cluster identification activity. If the same clusters are identified by means of the two algorithms, these can be considered valid. On the contrary, if clusters identified by the PCA are not observable using the method of random projections (in the same number of dimensions), a note of caution is required and the significance of the clusters must be evaluated in the light of biochemical knowledge about the protein. In this way, the technique of random projection represents a simple and intuitive way to evaluate the result of the PCA-based clustering algorithms.

We obtain the same cluster of structure by both algorithms in the case of HSA, COX-2 and Hb, with the single exception of the COX-1 case. However, as recalled above, it is a well known fact that a single stable structure is the dominant conformation of the COX-2, which is extremely similar to COX-1. Although it is true that there must be other conformations in the catalytic cycle of the COX-2, they must be only transient. This makes to think that, in the case of this enzyme, allostery models without conformational changes should be seriously taken into account. In fact plausible models of allostery without conformational changes have been proposed some time ago [60]. Our results suggest that this could be also the case for the COX-2 enzyme. The obtained data also show that the random

projection can be a simple way to validate the data obtained by PCA in the presence of a number of data lower than the degrees of freedom of the system.

Sepsis induces changes in both protein synthesis and structure, independently from the general inflammatory response. The underlying inflammatory process takes place in order to neutralise the causative agents, also due to to various modifications of the metabolic asset and the generation of molecular isoforms of the biochemical pre- and newly formed mediators [61,62]. The availability of new tools for protein study may be perspectively useful to better understand such events and their possible implications for new diagnostic tests and more effective therapies.

4. Methods

Atomic coordinates of the selected proteins were obtained from PDB [2]. To obtain the data sets in a matrix form, the pdb files were loaded in VMD (Visual Molecular Dynamics) [63] and superposed by a simple Tcl (Tool command language) scrip (www.tcl.tk). The α-carbon atom coordinates were extracted from the updated pdb files and written in a text file such that each row described a structure by Tcl scripting. The raw text file were edited by Vim (Vi IMproved) scripting (www.vim.org), so as to obtain the data matrix in a readable file format by the numerical analysis software (see below).

When we are dealing with protein structure datasets, the correlation matrix (henceforth indicated as *C*) should be obtained from the Cartesian coordinates of the atoms included in the analysis that represent the degrees of freedom of the system (also covariance matrix could be used). In its classical implementation, the normalized PCA is based on the eigenvector decomposition of the correlation matrix [54,55,58,64–66]. After the centroid subtraction, the covariance matrix of the data set can be obtained as

$$\chi_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

where $\langle \ldots \rangle$ represents the average over all the conformations in the data set. Then the correlation matrix is calculated from the *C*-matrix as

$$P_{ij} = \frac{\chi_{ij}}{\sqrt{\chi_{ii}\chi_{jj}}}$$

and this square symmetric matrix is diagonalized as

$$R^T P R = \Lambda$$

using standard numerical routines, where *R* is an orthonormal transformation matrix, the superscript ^{T} means transposition and Λ is a diagonal matrix whose elements are the eigenvalues. The eigenvalues, and the corresponding eigenvectors, are ordered in descending order of the eigenvalues. The empirical matrix was projected onto the eigenvectors to give the so called principal components. To overcome the limitations imposed by the number of replicas required for the correct evaluation of the covariance matrix, algorithms have been proposed, able to estimate the principal components also in the case of not well dimensioned data sets [14,67]. However, PCA is not the only algorithm that can perform the dimensionality reduction and the related unsupervised clustering tasks. A new and promising class of unsupervised learning algorithms [18,68–72] is represented by those that use some random projection methods. Correlation matrices of the protein structures obtained from molecular dynamics experiments [54–56,58,59] exhibit spectra whose bulk eigenvalues can be modeled by some symmetric random matrices [54–57], suggesting that a random matrix [54–57] can be used to obtain a system on which to project the data set [18]. The random projection algorithm that will be used here [18] works exactly as PCA, with the only difference that the matrix *C* is replaced by a symmetric random matrix of the same dimension of *C*. This relax the minimum number of samples required for the analysis of data sets containing a large number of degrees of freedom, making then analyzable also small crystallographic data sets, in which the number of different structures is much smaller than the degrees of freedom required to describe a protein.

PCA and random projection algorithms were implemented numerically in the Python language (www.python.org) in an IPython notebook [73]. The NumPy numerical software library [74] was used, which is part of the Scipy [75] software package. Matplotlib [76] package was used to obtain the all graphical outputs (obtained from Scipy; www.scipy.org). Before proceeding with the analysis of data, a preprocessing step that can be described as

$$x_{std}^i = \frac{x^i - \mu_x}{\sigma_x}$$

was applied [66], where μ_x is the sample mean of a particular degree of freedom column and σ_x the corresponding standard deviation, using the appropriate scikit-learn [77] built-in function. For PCA, the truncated SVD algorithm implemented in the scikit-learn software package was used [14,77]. The random projection algorithm and its practical implementation has been described in details elsewhere [18].

The confidence ellipses have been calculated assuming the normal distribution for the projected data and considering that the sum of the squares of Gaussian data is described by the Chi-square distribution.

References

1. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide protein data bank. *Nat. Struct. Biol.* **2003**, *10*, 980. [CrossRef] [PubMed]
2. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]
3. Rose, P.W.; Prlić, A.; Altunkaya, A.; Bi, C.; Bradley, A.R.; Christie, C.H.; Di Costanzo, L.; Duarte, J.M.; Dutta, S.; Feng, Z.; et al. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2017**, *45*, D271–D281. [PubMed]
4. Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality reduction: A comparative analysis. *J. Mach. Learn. Res.* **2009**, *10*, 66–71.
5. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559–572. [CrossRef]
6. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417. [CrossRef]
7. Ringnér, M. What is principal component analysis? *Nat. biotechnol.* **2008**, *26*, 303–304. [CrossRef]
8. Caruso, G.; Gattone, S.A.; Balzanella, A.; Di Battista, T. Cluster analysis: An application to a real mixed-type data set. In *Models and Theories in Social Systems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 525–533.
9. Caruso, G.; Gattone, S.A.; Fortuna, F.; Di Battista, T. Cluster analysis as a decision-making tool: A methodological review. In *International Symposium on Distributed Computing and Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 48–55.
10. Di Battista, T.; Fortuna, F. Clustering dichotomously scored items through functional k-means algorithm. *Electron. J. Appl. Stat. Anal.* **2016**, *9*, 433–450.
11. Gattone, S.A.; Giordani, P.; Di Battista, T.; Fortuna, F. Adaptive cluster double sampling with post stratification with application to an epiphytic lichen community. *Environ. Ecol. Stat.* **2018**, *25*, 125–138. [CrossRef]
12. Palese, L.L. Conformations of the HIV-1 protease: A crystal structure data set analysis. *Biochim. Biophys. Acta* **2017**, *1865*, 1416–1422. [CrossRef]

13. Palese, L.L. Analysis of the conformations of the HIV-1 protease from a large crystallographic data set. *Data Brief* **2017**, *15*, 696–700. [CrossRef] [PubMed]

14. Halko, N.; Martinsson, P.G.; Tropp, J.A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **2011**, *53*, 217–288. [CrossRef]

15. Maida, I.; Zanna, P.; Guida, S.; Ferretta, A.; Cocco, T.; Palese, L.L.; Londei, P.; Benelli, D.; Azzariti, A.; Tommasi, S.; et al. Translational control mechanisms in cutaneous malignant melanoma: The role of eIF2α. *J. Transl. Med.* **2019**, *17*, 20. [CrossRef] [PubMed]

16. Palese, L.L. Cytochrome c oxidase structures suggest a four-state stochastic pump mechanism. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4822–4830. [CrossRef] [PubMed]

17. Acharya, K.R.; Lloyd, M.D. The advantages and limitations of protein crystal structures. *Trends Pharmacol. Sci.* **2005**, *26*, 10–14. [CrossRef] [PubMed]

18. Palese, L.L. A random version of principal component analysis in data clustering. *Comput. Biol. Chem.* **2018**, *73*, 57–64. [CrossRef]

19. Fanali, G.; di Masi, A.; Trezza, V.; Marino, M.; Fasano, M.; Ascenzi, P. Human serum albumin: From bench to bedside. *Mol. Aspects Med.* **2012**, *33*, 209–290. [CrossRef]

20. Nicholson, J.; Wolmarans, M.; Park, G. The role of albumin in critical illness. *Br. J. Anaesth.* **2000**, *85*, 599–610. [CrossRef]

21. Taverna, M.; Marie, A.L.; Mira, J.P.; Guidet, B. Specific antioxidant properties of human serum albumin. *Ann. Intensive Care* **2013**, *3*, 4. [CrossRef]

22. Ascenzi, P.; Fasano, M. Allostery in a monomeric protein: The case of human serum albumin. *Biophys. Chem.* **2010**, *148*, 16–22. [CrossRef]

23. Inchingolo, F.; Dipalma, G.; Cirulli, N.; Cantore, S.; Saini, R.; Altini, V.; Santacroce, L.; Ballini, A.; Saini, R. Microbiological results of improvement in periodontal condition by administration of oral probiotics. *J. Biol. Regul. Homeost. Agents* **2018**, *32*, 1323–1328. [PubMed]

24. Vinolo, M.; Rodrigues, H.; Nachbar, R.; Curi, R. Modulation of inflammatory and immune responses by short-chain fatty acids. In *Diet, Immunity and Inflammation*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 435–458.

25. Smith, W.L.; DeWitt, D.L.; Garavito, R.M. Cyclooxygenases: Structural, cellular, and molecular biology. *Annu. Rev. Biochem.* **2000**, *69*, 145–182. [CrossRef] [PubMed]

26. Garavito, R.M.; Malkowski, M.G.; DeWitt, D.L. The structures of prostaglandin endoperoxide H synthases-1 and-2. *Prostaglandins Other Lipid Mediat.* **2002**, *68*, 129–152. [CrossRef]

27. Smith, W.L.; Urade, Y.; Jakobsson, P.J. Enzymes of the cyclooxygenase pathways of prostanoid biosynthesis. *Chem. Rev.* **2011**, *111*, 5821–5865. [CrossRef] [PubMed]

28. Tunctan, B.; Korkmaz, B.; Nihal Sari, A.; Kacan, M.; Unsal, D.; Sami Serin, M.; Kemal Buharalioglu, C.; Sahan-Firat, S.; Schunck, W.H.; R Falck, J.; et al. A novel treatment strategy for sepsis and septic shock based on the interactions between prostanoids, nitric oxide, and 20-hydroxyeicosatetraenoic acid. *Antiinflamm. Antiallergy Agents Med. Chem.* **2012**, *11*, 121–150.

29. Dong, L.; Vecchio, A.J.; Sharma, N.P.; Jurban, B.J.; Malkowski, M.G.; Smith, W.L. Human cyclooxygenase-2 is a sequence homodimer that functions as a conformational heterodimer. *J. Biol. Chem.* **2011**, *286*, 19035–19046. [CrossRef]

30. Zou, H.; Yuan, C.; Dong, L.; Sidhu, R.S.; Hong, Y.H.; Kuklev, D.V.; Smith, W.L. Human cyclooxygenase-1 activity and its responses to COX inhibitors are allosterically regulated by nonsubstrate fatty acids. *J. Lipid Res.* **2012**, *53*, 1336–1347. [CrossRef]

31. Mitchener, M.M.; Hermanson, D.J.; Shockley, E.M.; Brown, H.A.; Lindsley, C.W.; Reese, J.; Rouzer, C.A.; Lopez, C.F.; Marnett, L.J. Competition and allostery govern substrate selectivity of cyclooxygenase-2. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12366–12371. [CrossRef]

32. Malkowski, M.G.; Thuresson, E.D.; Lakkides, K.M.; Rieke, C.J.; Micielli, R.; Smith, W.L.; Garavito, R.M. Structure of eicosapentaenoic and linoleic acids in the cyclooxygenase site of prostaglandin endoperoxide H synthase-1. *J. Biol. Chem.* **2001**, *276*, 37547–37555. [CrossRef]

33. Thuresson, E.D.; Malkowski, M.G.; Lakkides, K.M.; Rieke, C.J.; Mulichak, A.M.; Ginell, S.L.; Garavito, R.M.; Smith, W.L. Mutational and x-ray crystallographic analysis of the interaction of dihomo-γ-linolenic acid with prostaglandin endoperoxide H synthases. *J. Biol. Chem.* **2001**, *276*, 10358–10365. [CrossRef]

34. Malkowski, M.; Ginell, S.; Smith, W.; Garavito, R. The productive conformation of arachidonic acid bound to prostaglandin synthase. *Science* **2000**, *289*, 1933–1937. [CrossRef] [PubMed]

35. Harman, C.A.; Rieke, C.J.; Garavito, R.M.; Smith, W.L. Crystal structure of arachidonic acid bound to a mutant of prostaglandin endoperoxide H synthase-1 that forms predominantly 11-hydroperoxyeicosatetraenoic acid. *J. Biol. Chem.* **2004**, *279*, 42929–42935. [CrossRef] [PubMed]

36. Loll, P.J.; Picot, D.; Ekabo, O.; Garavito, R.M. Synthesis and Use of Iodinated Nonsteroidal Antiinflammatory Drug Analogs as Crystallographic Probes of the Prostaglandin H2 Synthase Cyclooxygenase Active Site. *Biochemistry* **1996**, *35*, 7330–7340. [CrossRef] [PubMed]

37. Yuan, Y.; Tam, M.F.; Simplaceanu, V.; Ho, C. New look at hemoglobin allostery. *Chem. Rev.* **2015**, *115*, 1702–1724. [CrossRef]

38. Brunori, M. Half a Century of Hemoglobin's Allostery. *Biophys. J.* **2015**, *109*, 1077–1079. [CrossRef]

39. Effenberger-Neidnicht, K.; Hartmann, M. Mechanisms of hemolysis during sepsis. *Inflammation* **2018**, *41*, 1569–1581. [CrossRef]

40. Bateman, R.; Sharpe, M.; Singer, M.; Ellis, C. The effect of sepsis on the erythrocyte. *Int. J. Mol. Sci.* **2017**, *18*, 1932. [CrossRef]

41. Santacroce, L.; Losacco, T. Abdominal sepsis in surgical patients. Pathophysiology and prevention. *Recenti Prog. Med.* **2006**, *97*, 411–416.

42. Yoo, H.; Ku, S.K.; Kim, S.W.; Bae, J.S. Early diagnosis of sepsis using serum hemoglobin subunit Beta. *Inflammation* **2015**, *38*, 394–399. [CrossRef]

43. Jiang, Y.; Jiang, F.Q.; Kong, F.; An, M.M.; Jin, B.B.; Cao, D.; Gong, P. Inflammatory anemia-associated parameters are related to 28-day mortality in patients with sepsis admitted to the ICU: A preliminary observational study. *Ann. Intensive Care* **2019**, *9*, 67. [CrossRef]

44. Perutz, M.F.; Rossmann, M.G.; Cullis, A.F.; Muirhead, H.; Will, G.; North, A. Structure of haemoglobin: A three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* **1960**, *185*, 416–422. [CrossRef] [PubMed]

45. Monod, J.; Wyman, J.; Changeux, J.P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88–118. [CrossRef]

46. Perutz, M. Stereochemistry of cooperative effects in haemoglobin: Haem–haem interaction and the problem of allostery. *Nature* **1970**, *228*, 726–734. [CrossRef] [PubMed]

47. Baldwin, J.; Chothia, C. Haemoglobin: The structural changes related to ligand binding and its allosteric mechanism. *J. Mol. Biol.* **1979**, *129*, 175–220. [CrossRef]

48. Dey, S.; Chakrabarti, P.; Janin, J. A survey of hemoglobin quaternary structures. *Proteins* **2011**, *79*, 2861–2870. [CrossRef] [PubMed]

49. Miyazaki, G.; Morimoto, H.; Yun, K.M.; Park, S.Y.; Nakagawa, A.; Minagawa, H.; Shibayama, N. Magnesium (II) and zinc (II)-protoporphyrin IX's stabilize the lowest oxygen affinity state of human hemoglobin even more strongly than deoxyheme. *J. Mol. Biol.* **1999**, *292*, 1121–1136. [CrossRef] [PubMed]

50. Waller, D.; Liddington, R. Refinement of a partially oxygenated T state human haemoglobin at 1.5 Å resolution. *Acta Crystallogr. B* **1990**, *46*, 409–418. [CrossRef]

51. Kavanaugh, J.S.; Rogers, P.H.; Arnone, A. Crystallographic Evidence for a New Ensemble of Ligand-Induced Allosteric Transitions in Hemoglobin: The T-to-T$_{High}$ Quaternary Transitions. *Biochemistry* **2005**, *44*, 6101–6121. [CrossRef]

52. Sen, U.; Dasgupta, J.; Choudhury, D.; Datta, P.; Chakrabarti, A.; Chakrabarty, S.B.; Chakrabarty, A.; Dattagupta, J.K. Crystal structures of HbA$_2$ and HbE and modeling of hemoglobin δ4: interpretation of the thermal stability and the antisickling effect of HbA$_2$ and identification of the ferrocyanide binding site in Hb. *Biochemistry* **2004**, *43*, 12477–12488. [CrossRef]

53. Schumacher, M.A.; Dixon, M.M.; Kluger, R.; Jones, R.T.; Brennan, R.G. Allosteric transition intermediates modeled by cross-linked hemoglobins. *Nature* **1995**, *375*, 84–87. [CrossRef]

54. Palese, L.L. Random Matrix Theory in molecular dynamics analysis. *Biophys. Chem.* **2015**, *196*, 1–9. [CrossRef] [PubMed]

55. Palese, L.L. Correlation Analysis of Trp-Cage Dynamics in Folded and Unfolded States. *J. Phys. Chem. B* **2015**, *119*, 15568–15573. [CrossRef] [PubMed]

56. Palese, L.L. Protein States as Symmetry Transitions in the Correlation Matrices. *J. Phys. Chem. B* **2016**, *120*, 11428–11435. [CrossRef] [PubMed]

57. Edelman, A.; Wang, Y. Random matrix theory and its innovative applications. In *Advances in Applied Mathematics, Modeling, and Computational Science*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 91–116.

58. Bossis, F.; Palese, L.L. Amyloid beta (1–42) in aqueous environments: Effects of ionic strength and E22Q (Dutch) mutation. *Biochim. Biophys. Acta* **2013**, *1834*, 2486–2493. [CrossRef] [PubMed]

59. Palese, L.L. Protein dynamics: Complex by itself. *Complexity* **2013**, *18*, 48–56. [CrossRef]

60. Cooper, A.; Dryden, D. Allostery without conformational change. *Eur. Biophys. J.* **1984**, *11*, 103–109. [CrossRef] [PubMed]

61. Trentadue, R.; Fiore, F.; Massaro, F.; Papa, F.; Iuso, A.; Scacco, S.; Santacroce, L.; Brienza, N. Induction of mitochondrial dysfunction and oxidative stress in human fibroblast cultures exposed to serum from septic patients. *Life Sci.* **2012**, *91*, 237.

62. Bosmann, M.; Ward, P.A. The inflammatory response in sepsis. *Trends Immunol.* **2013**, *34*, 129–136. [CrossRef]

63. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

64. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [CrossRef]

65. Shlens, J. A tutorial on principal component analysis. *arXiv* **2014**, arXiv:1404.1100.

66. Raschka, S. *Python Machine Learning*; Packt Publishing: Birmingham, UK, 2015.

67. Roweis, S. EM algorithms for PCA and SPCA. In Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems, Denver, CO, USA, 29 November–4 December 1998; pp. 626–632.

68. Johnson, W.B.; Lindenstrauss, J. Extensions of Lipschitz mappings into a Hilbert space. *Cont. Math.* **1984**, *26*, 1.

69. Papadimitriou, C.H.; Tamaki, H.; Raghavan, P.; Vempala, S. Latent semantic indexing: A probabilistic analysis. *J. Comput. Syst. Sci.* **1998**, *61*, 159–168.

70. Kaski, S. Dimensionality reduction by random mapping: Fast similarity computation for clustering. In Proceedings of the 1998 IEEE World Congress on Computational Intelligence, Anchorage, Alaska, 4–9 May 1998; Volume 1, pp. 413–418.

71. Achlioptas, D. Database-friendly random projections. In Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Santa Barbara, CA, USA, 21–23 May 2001; pp. 274–281.

72. Bingham, E.; Mannila, H. Random projection in dimensionality reduction: Applications to image and text data. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 26–29 August 2001; pp. 245–250.

73. Pérez, F.; Granger, B.E. IPython: A system for interactive scientific computing. *Comput. Sci. Eng.* **2007**, *9*, 21–29. [CrossRef]

74. Van Der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [CrossRef]

75. Oliphant, T.E. Python for scientific computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20. [CrossRef]

76. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]

77. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.