JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

pubs.acs.org/JCTC

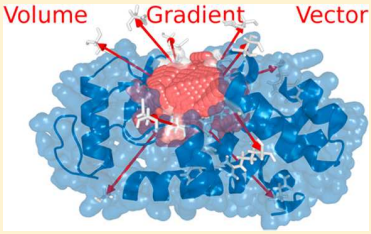
Protein Fluctuations and Cavity Changes Relationship

German P. Barletta and Sebastian Fernandez-Alberti*

Universidad Nacional de Quilmes/CONICET, Roque Saenz Peña 352, B1876BXD Bernal, Argentina

Supporting Information

ABSTRACT: Protein cavities and tunnels are critical for function. Ligand recognition and binding, transport, and enzyme catalysis require cavities rearrangements. Therefore, the flexibility of cavities should be guaranteed by protein vibrational dynamics. Molecular dynamics simulations provide a framework to explore conformational plasticity of protein cavities. Herein, we present a novel procedure to characterize the dynamics of protein cavities in terms of their volume gradient vector. For this purpose, we make use of algorithms for calculation of the cavity volume that result robust for numerical differentiations. Volume gradient vector is expressed in terms of principal component analysis obtained from equilibrated molecular dynamics simulations. We analyze contributions of principal component modes to the volume gradient vector according to their frequency and degree of delocalization. In all our test cases, we find that low frequency modes play a critical role together with minor contributions of high frequency modes. These modes involve concerted motions of significant fractions of the total residues lining the cavities. We make use of variations of the potential energy of a protein in the direction of the volume gradient vector as a measure of flexibility of the cavity. We show that proteins whose collective low frequency fluctuations contribute the most to changes of cavity volume exhibit more flexible cavities.



I. INTRODUCTION

Protein cavities and tunnels are structural features that play key roles in biological functions. They are involved in a large variety of processes like ligand recognition and binding, transport and/or transfer of small molecules, and also intra- and intermolecular energy transfer subject to structural reorganizations.[1] Therefore, they are the subject of research for functional assignment and drug design.[2−4] Cavities represent not only empty or water-containing spaces that can potentially contribute to protein plasticity but create specific microenvironments required for concerted interactions that participate in enzyme catalysis. The connection between structures of cavities and the native state of proteins, represented by an ensemble of conformers in dynamics equilibrium, has been extensively studied.[5−7] Either rotations of a single residue side chain or large collective interdomain conformational changes can control the transit of substrates and products within the protein.[8,9] Protein fluctuations and their relationship with cavity changes are essential for a complete description of protein function.

Properties of protein cavities, like size, polarity, solvation, and residue type constituency, have been investigated in detail for different kind of them: within domains, between domains, and between protein subunits.[10−12] This knowledge provides useful guidelines for protein modeling and design.[13]

Protein cavities are not static objects.[14−16] The flexibility and dynamics of cavities are widely acknowledged to be crucial for function.[17] Conformational changes frequently involve changes in cavity volumes with subsequent impacts on ligand affinity and specificity and, therefore, regulation of the biological function of a protein.[18,19] While detailed structural description of cavities is useful to predict the type of ligands that interact with the protein, their dynamics features contribute to analyze potentials degrees of binding promiscuity.[20,21] Furthermore, the shape, size, and flexibility of cavities have been recently associated with the degree of conformational diversity and disorder regions, among others structural and dynamical features of proteins.[22]

Flexibility of cavities is provided by protein vibrational dynamics ranging from slow and collective motions to fast and localized motions. Molecular dynamics (MD) simulations[23−26] combined with principal component analysis (PCA)[27−31] provide a framework for decomposing the complexity of proteins motions into decoupled individual contributions. This combination of MD and PCA has recently been applied to develop a procedure that reveals the existence of correlations between the dynamics of cavities and structures.[32]

A large variety of methods have been developed in order to achieve accurate values of cavity volumes. Usually, the protein is represented as a set of spheres with volumes related to atomic van der Waals radius. One type of methods considers the protein embedded in a three-dimensional grid of voxels followed by subsequent counting of unoccupied ones.[33−40] Another type of method implies the use of spherical probes to fill the cavity.[41−43,40] A third type of methods is based on geometric algorithms that make use of three-dimensional triangulations of atomic positions, such as α-shape, Delaunay, and Voronoi algorithms, among others.[44−49,8]

In this paper, we present a novel procedure to analyze the dynamics of protein cavities. We characterize it in terms of their volume gradient vector ($\nabla \mathbf{V}_{ol}$). For this purpose, we make use

Received: July 12, 2017
Published: December 20, 2017

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

of algorithms for calculation of cavity volumes that result robust for numerical differentiations. $\nabla \mathbf{V}_{ol}$ is expressed in terms of PCA modes. As a result, dynamics features of cavities can be straightforward connected to thermal fluctuations in proteins.

The paper is organized as follows. Theoretical methods are described in section II. Our results are presented and discussed in section III. Finally, section IV summarizes our findings and conclusions.

II. METHODS

A. Cases of Study. The human immunodeficiency virus type 1 protease (HIVP-1) is a homodimer that presents a cavity with an aperture involving low-frequency breathing motions in which participate synchronized movements of rigid backbone segments.[1,50] The active site is symmetrically formed by both monomers. The access to the active site is controlled by two β-turn flaps with specific anchoring regions. HIVP-1 protease can be divided into three domains: the core domain, localized at the interface containing the active site, the terminal domain including both N and C terminals, and the flap domain, the most flexible domain[51−53] with the two β-turns that act as a gate. HIVP-1 is a good example of a cavity acting as a gate whose opening/closing motions are dependent on hinge and anchoring regions. Therefore, mutations on specific residues lead to a significant impact on the exposure of the active site. This is the case of significant changes analyzed on the gated association rate constants with different drugs of F53L and G48V/V82A/I84V/L90M mutants among others.[50]

The epidermal growth factor receptor (EGFR) kinase presents a main active site pocket limited by an N-terminal and C-terminal lobes (N and C lobes) that are connected by a hinge region.[54] While the N lobe is composed of five β-strands and the αC helix, the C lobe is predominantly helical containing a highly flexible activation loop. Two main conformations, so-called active and inactive conformers[55,56] have been largely structurally described, and their relative stabilities have been associated with different types of cancers.[57,58] Some structural characteristics have been proposed to differentiate kinase conformations, but these considerations could lead to ambiguous classifications. Despite the fact that common structural characteristics shared among the main pockets of most of active EGFR kinase mutants have been proposed, several reported controversies and ambiguous conformation classifications can be found.[59] Furthermore, different single amino acid substitutions (SASs) have been proven to alter the equilibrium of pre-existing active/inactive conformer populations. In this context, the identification of residues whose vibrational motions participate the most in changes in the size and shape of the active site pocket of the EGFR kinase active conformers contributes to clarify certain SASs effects.

The T4 lysozyme (T4L) has been largely studied by either experimental[60−64] and theoretical techniques.[65−69] The volume of the catalytic cleft is functionally significant since the opening and closure of the cleft is associated with ligand binding.[70] Previous PCA studies[65,70] show that its first PCA modes are associated with N-terminal and C-terminal interdomain motions: hinge-bending, twisting, and relative torsions. These motions have a substantial impact on the size and shape of the cavity. Within this context, T4L represents a good example of significant contributions of delocalized functional low-frequency modes dominating the closing and opening of the catalytic cleft, presumably facilitating substrate binding and release.

B. MD Simulations and PCA. MD simulations are performed for each case of study (see section II.A). AMBER 16 software package[71,72] is used for all simulations. Initial coordinates for each protein are taken from X-ray structures with Protein Data Bank[73] entries: 1HVR (HIV protease), 1M14 (EGFR active conformer), and 256L (T4 Lysozyme). Ions are added for charge neutralization. Each system is solvated with explicit TIP3P[74] water molecules in a truncated octahedric periodic box large enough to contain the protein and 10 Å of solvent on all sides. In all the MD calculations, the all-hydrogen topology with Amber ff14SB[75,76] force field parameters is used, except for 2 *S*-hydroxycysteine amino acids and the XK2 ligand of HIV protease where GAFF (General Amber Force Field)[77,78] is employed using AM1-BCC charges.[79] Minimization of each molecular system is performed by 100-step of steepest-descent and 400-step conjugate gradient minimizations applying constraints to the protein atoms. These are followed by a 400-step unconstrained conjugate gradient minimization. The systems are then heated for 150 ps until they reach the final temperature of 300 K. During heating, a harmonic restraint of 50.0 kcal/(mol·Å^2) is applied to the protein atoms. In all simulations, a 2 fs time step with SHAKE algorithm applied to all bonds involving hydrogen is used. Periodic boundary conditions and particle-mesh Ewald (PME) sums are used and a cutoff of 10 Å is applied to nonbonded interactions. The systems are equilibrated at constant pressure using 26 steps of 100 ps and reducing the restraint on each step. After the last step with restraints, all restraints are lifted and final equilibrations are achieved by 300 ns simulations at the constant temperature of 300 K using Andersen barostat and Langevin thermostat with a γ collision frequency of 2 ps^{-1}. Finally, 600 ns MD simulations are performed in order to collect equilibrated configurations at 10 ps intervals.

PCA[27,28,80,30,31] is performed from previous K collected configurations by calculating the covariance matrix $\mathbf{C}$ with elements

$$C_{ij} = \langle q_i q_j \rangle = \frac{1}{K} \sum_{k=1}^{K} q_i^k q_j^k \tag{1}$$

with $q_i^k = \sqrt{\langle m_i \rangle}\,(x_i^k - x_i)$ being the mass-weighted internal displacement of Cartesian coordinate x_i^k of the ith atom with mass m_i, and the angular brackets represent the average obtained from the K equilibrated configurations, with the center-of-mass translational and rotational motions previously being removed by least-squares fitting of a reference structure.[81]

PCA modes are obtained by diagonalizing the $\mathbf{C}$ matrix as

$$\mathbf{Q}^{\mathrm{T}}\mathbf{C}\mathbf{Q} = \mathbf{\Lambda} \tag{2}$$

where $\mathbf{Q}$ is the eigenvector matrix whose columns $\mathbf{Q}_i$ are the PCA modes with frequencies given as $\frac{1}{\sqrt{\lambda_k}}$, being λ_k the elements of the eigenvalue matrix $\mathbf{\Lambda}$. In the present work, PCA modes are calculated considering only the C_α atoms in the proteins, so $\mathbf{Q}_i$ are vectors of dimension $3N$, where N is the number of residues of the molecule. Any $\mathbf{Q}_i$ is expressed in terms of mass-weighted internal displacements Cartesian coordinates as

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\mathbf{Q}_i = \sum_{j=1}^{N} \sum_{l=1}^{3} Q_{ji}^{l} \hat{\mathbf{q}}_j^{l} \tag{3}$$

where $\hat{\mathbf{q}}_j^l$ is a unit vector in the direction of the l coordinate (l = 1, 2, 3 for x, y, z, respectively) of the mass-weighted internal displacement Cartesian vector for the jth residue.

C. Cavity Volume. $\nabla \mathbf{V}_{\mathbf{ol}}$ corresponding to a protein cavity is evaluated approximating derivatives by finite differences. Therefore, we make use of a combination of algorithms that result robust for numerical differentiations: convex hull algorithm to identify a specific inclusion area, Delaunay triangulation to evaluate the initial cavity volume and displacements (non-Delaunay) of the resulting tetrahedrons vertices to track the volume change.

C.1. Convex Hull. To determine the cavity of each protein, an inclusion area is defined using the convex hull method.[82] The coordinates of a predefined set of atoms surrounding a cavity represents the set of points that define the convex hull as the smallest convex set that contains them. The convex hull can be visualized as the shape enclosed by a rubber band stretched around them (see Figure 1a).

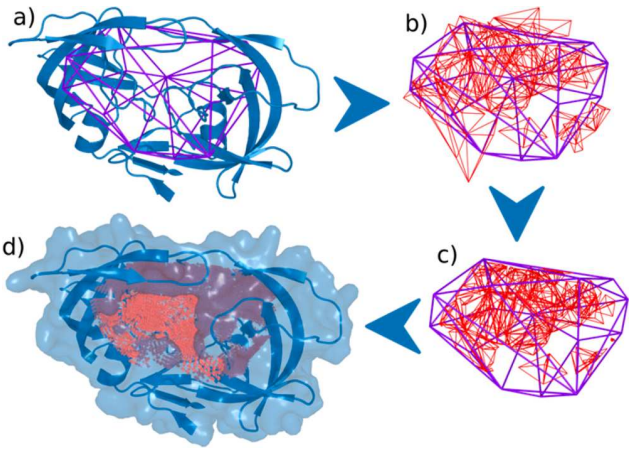


Figure 1. Scheme of the algorithm used for cavity volume calculation: (a) convex hull; (b) Delaunay triangulation; (c) refinement; (d) final cavity volume.

C.2. Delaunay Triangulation. A Delaunay triangulation of a set of points (i.e., atoms surrounding the cavity) is a triangulation such that no point lies inside the circumcircle of any triangle. Delaunay triangulations maximize the minimum angle of all the angles of the triangles in the triangulation, trying to avoid sliver triangles. Triangulation is shown in Figure 1b, where triangles tile the convex hull area. As a result, a set of tetrahedrons with atoms that define the cavity as its vertices is obtained. Since tetrahedrons satisfy the Delaunay condition, i.e., the requirement that the circumcircles of all triangles have empty interiors, the volumes of the resulting tetrahedrons can be used as a first step in the calculation of the cavity volume. Only tetrahedrons with at least one vertex inside the convex hull are retained. Besides, considering that atoms are not points and occupy a certain volume, tetrahedrons smaller than 1.3 Å are discarded.

C.3. Refinement. Since some tetrahedrons intersect the inclusion area but are not entirely contained in it, the intersections between the tetrahedrons and the inclusion area are calculated, and only the volume that is inside the inclusion area is kept. The result of this process is displayed in Figure 1c. Finally, the effective volume of each atom, considered as a sphere of radius equal to its corresponding van der Waals radius, is subtracted to obtain the final volume of the cavity (see Figure 1d).

D. Volume Gradient Vector ($\nabla \mathbf{V}_{\mathbf{ol}}$). The $\nabla \mathbf{V}_{\mathbf{ol}}$ of a protein cavity, previously defined following the procedure denoted in section II.C, is obtained by partial derivatives of the cavity volume in the basis of PCA modes $\{\mathbf{Q}_i\}_{i=1,3N}$, being N the number of protein residues

$$\nabla \mathbf{V}_{\mathbf{ol}} = \sum_{i=1}^{3N} c_i \mathbf{Q}_i = \sum_{i=1}^{3N} \frac{\partial \mathbf{V}_{\mathbf{ol}}}{\partial \mathbf{Q}_i} \mathbf{Q}_i \tag{4}$$

with $\frac{\partial \mathbf{V}_{\mathbf{ol}}}{\partial \mathbf{Q}_i}$ being the partial derivative of the cavity volume respect to the ith PCA mode calculated by finite differences

$$\frac{\partial \mathbf{V}_{\mathbf{ol}}}{\partial \mathbf{Q}_i} = \frac{\mathbf{V}_{\mathbf{ol}} - \mathbf{V}_{\mathbf{ol,eq}}}{\Delta \mathbf{Q}_i} \tag{5}$$

being $\mathbf{V}_{\mathbf{ol}}$ and $\mathbf{V}_{\mathbf{ol,eq}}$ the cavity volumes of distorted structure in the direction of $\mathbf{Q}_i$ and reference average structure, respectively. $\Delta \mathbf{Q}_i = \mathbf{Q}_i - \mathbf{Q}_{i,\mathrm{eq}}$ is the relative displacement of coordinate Q_i, that is considered as a fixed fraction for all modes of to the corresponding thermal amplitudes $A_i = (2\lambda_i)^{1/2}$. Since PCA modes are calculated on C_α atoms, all the atoms in a residue are displaced along each PCA mode as a rigid block. In order to avoid steric hindrances between atoms due to displacements performed during calculation of partial derivatives by finite differences, structural distortions were restricted below a root-mean square difference (RMSD) of 0.05 Å.

After a Delaunay triangulation is performed, many small tetrahedrons are formed between neighboring atoms and even though there are no atom centers inside them, these tetrahedrons are filled with the electronic density of the atoms at its vertices. These tetrahedrons do not contribute to the final cavity volume and slight displacements of their vertices (atoms), due to small structural distortions, can lead to violations of the Delaunay condition. This can force to an undesired entire retriangulation during numerical calculations of partial derivatives. To avoid this issue, an initial triangulation was performed on the average structure of the MD run. Only relevant tetrahedrons were kept, neglecting small tetrahedrons outlined above. The vertices (atoms) of the remained tetrahedrons are then displaced along the PCA modes. In this way, volume changes are calculated without updating the triangulation, avoiding any "jump" in the volume magnitude and ensuring a linear response to each PCA mode.

III. RESULTS AND DISCUSSION

Herein, four cases of study are presented (see section II.A) in order to enlighten general dynamics aspects of cavities and their connections with protein collective vibrational motions. We do not attempt to perform detailed analysis of each of these systems but rather present complementary dynamics aspects provided by the analysis of $\nabla \mathbf{V}_{\mathbf{ol}}$. While there is no consensus about the definition of cavities,[83] herein we define them according to visual analysis[84] and previous knowledge on each system. In the Supporting Information (Table S1), we provide the complete list of residues that define molecular surfaces enclosing the inner cavity volumes for each of the cases of study.

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

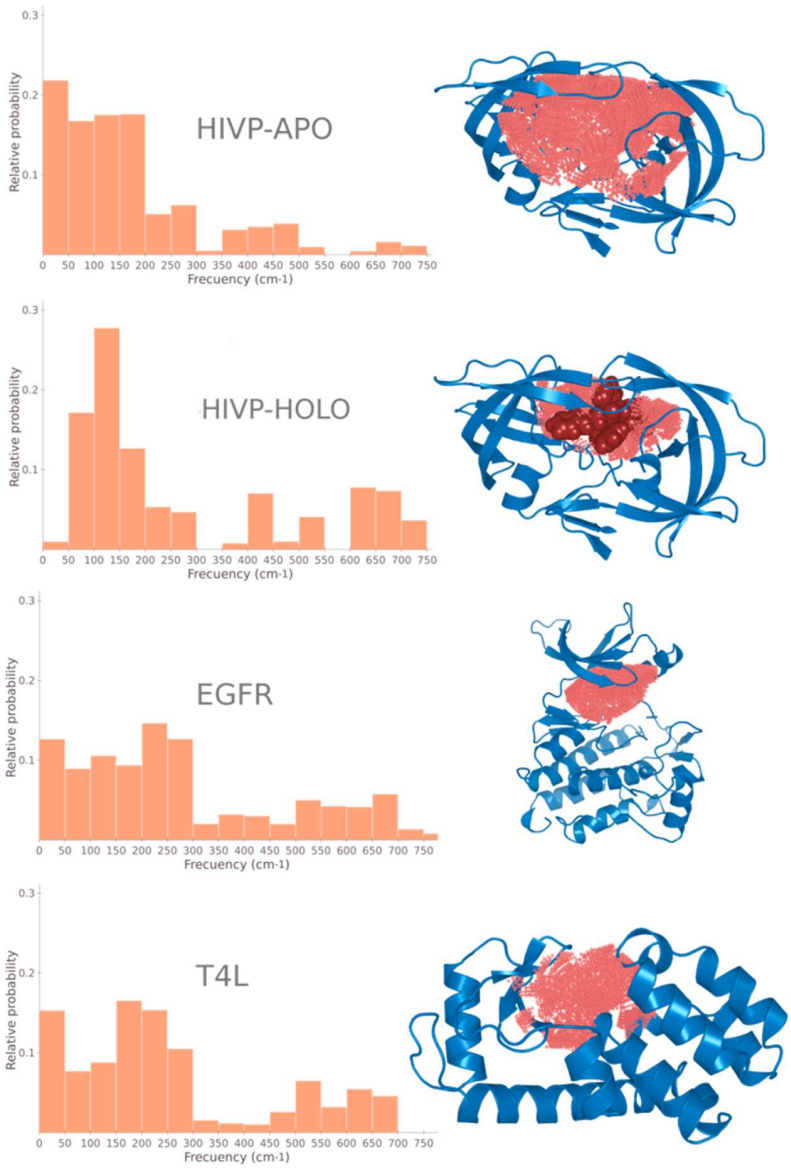


Figure 2. (left) Distribution of contributions of PCA modes to the direction of maximum change of the cavity volumes ($\nabla \mathbf{V}_{ol}$) as a function of mode frequencies obtained for each of our cases of study: HIVP-APO, HIVP-HOLO, EGFR, and T4L. (right) Corresponding structures represented by secondary structure motifs, where selected cavities are denoted in red, and proteins, blue.

The effect of vibrational motions on protein cavities can be analyzed developing $\nabla \mathbf{V}_{ol}$ in terms of PCA modes (see eq 4). In order to stress differences among vibrations that contribute the most to cavity changes, we retain the first P vibrations ordered by decreasing values of c_i^2, defining the participation number P as

$$P_q = \left(\sum_{i=1}^{3N} c_i^4 \right)^{-1} \tag{6}$$

P_q represents a measure of the delocalization of $\nabla \mathbf{V}_{ol}$ on the basis of PCA modes. Its value, rounded to the nearest higher integer, indicates the minimum number of modes needed to represent $\nabla \mathbf{V}_{ol}$. While values of $P_q \approx 3N$ mean that $\nabla \mathbf{V}_{ol}$ is expanded among all PCA modes, $P_q \approx 1$ indicate that one single mode dominates the $\nabla \mathbf{V}_{ol}$ direction. The first P_q modes ordered by decreasing values of c_i^2 are the minimum set of modes required to achieve a good description of $\nabla \mathbf{V}_{ol}$ neglecting minor contributions and, therefore, stressing the main *protein fluctuations−cavity changes* relationships. Values of $P_q/3N$ = 0.25, 0.16, 0.20, and 0.26 were obtained for HIVP-APO, HIVP-HOLO, EGFR, and T4L, respectively. That is, only a small fraction of PCA modes significantly contribute to cavity changes. From now on, our analysis is focused on these first P_q modes that contribute the most to $\nabla \mathbf{V}_{ol}$.

Figure 2 shows the distribution of contributions of PCA modes (c_i^2) to $\nabla \mathbf{V}_{ol}$ throughout the whole frequency range. For all systems, low-frequency modes, within the range up to 300 cm^{-1}, represent the main contributions to changes in the cavity volumes. Minor contributions come from high-frequency modes. It is interesting to note that, in the case of HIVP, ligand−protein interactions increase contributions of high frequency modes. This can be a consequence of the rigidization of the cavity that follows ligand binding, hindering the effect of collective low-frequency modes.

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

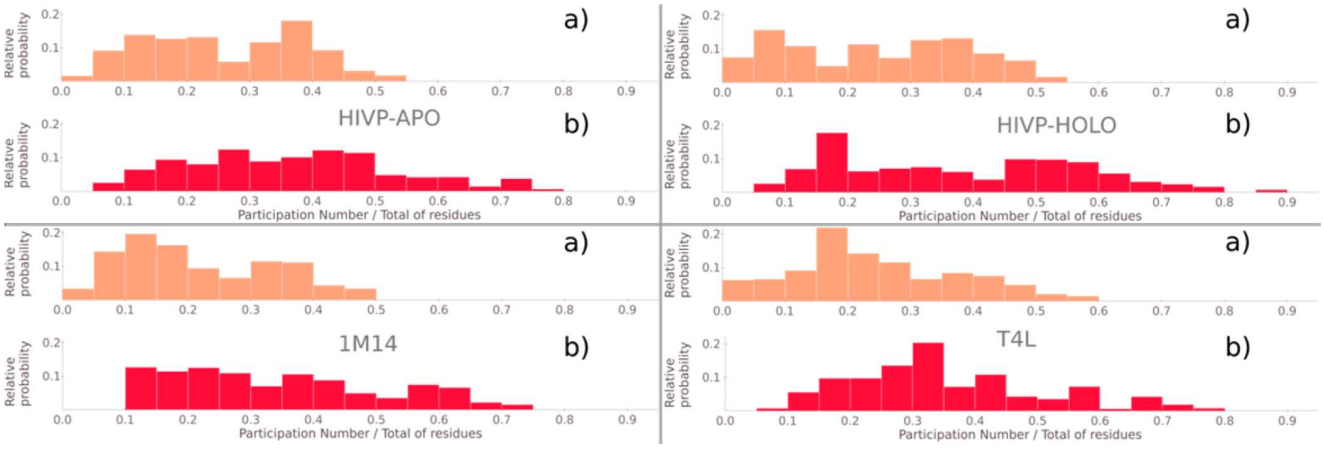


Figure 3. Distribution of the fraction of residues involved in the motion of PCA modes weighted by their contribution to the volume gradient $\nabla \mathbf{V}_{ol}$ considering (a) all residues of the protein and (b) only residues on the surface of the cavities.

It is interesting at this point to examine the convergence in our results. For this purpose, we have partitioned our simulations in two simulations of 300 ns each and new sets of PCA modes have been obtained for each half of simulation. Thereafter, we recalculate the distributions previously shown in Figure 2. Results for each pair of 300 ns simulations are shown as Supporting Information (Figure S1). In order to quantify the overlap between them, the Bhattacharyya coefficient (BCF) has been calculated.[85,86] BCF is a divergence-type measure between distributions that takes values of 1 for identical distributions and approaches to 0 as both distributions differ. Values of 0.97, 0.98, 0.97, and 0.97 have been obtained for BCF between each pair of histograms for HIVP-APO, HIVP-HOLO, EGFR, and T4L, respectively. Besides, the overlap values between pairs of volume gradient vectors $\nabla \mathbf{V}_{ol}$, obtained using PCA modes corresponding to each pair of 300 ns simulations of HIVP-APO, HIVP-HOLO, EGFR, and T4L were 0.70, 0.97, 0.82, and 0.91, respectively. These results confirm the convergence of conformational sampling for our MD simulations and the PCA results derived from them.

PCA modes can be analyzed in terms of their localization/delocalization throughout the protein. This can be done by calculating the residue participation number for each PCA mode defined by

$$P_i = \left(\sum_{j=1}^{N} {Q_{ji}}^4 \right)^{-1} \tag{7}$$

where ${Q_{ji}}^2 = ({Q_{ji}}^x)^2 + ({Q_{ji}}^y)^2 + ({Q_{ji}}^z)^2$ and ${Q_{ji}}^l = (l = x, y, z)$ are the components of the jth atom in the ith normal mode. Values of $P_i \approx N$ describe fluctuations equally distributed throughout all the residues of the protein, and $P_i \approx 1$ corresponds to fluctuations of a single residue. In Figure 3a we show the distribution of the fraction of residues involved in the motion of PCA modes weighted by their contribution to the volume gradient $\nabla \mathbf{V}_{ol}$. In all our cases of study, fluctuations that contribute to maximize the cavity volume can involve up to approximately half of the total residues of the protein. This is expected due to the main contributions of collective low-frequency modes to cavity changes previously shown in Figure 2. In order to further analyze this feature, we reduce our analysis to residues lining the cavities. For this purpose, PCA modes were reduced and renormalized retaining only elements involving these residues. Figure 3b displays the corresponding distribution of the fraction of residues on the surface of the cavities that participate together of concerted motions due to collective modes related to cavity changes. In all cases, up to 70% of residues of the cavities participate in collective fluctuations with a direct impact on the size of the cavities. That is, most of residues lining the cavities can be displaced in a synchronized way that maximizes effects on cavity sizes and shapes. These low-frequency structural modes, while involving movements of residues localized throughout the protein (i.e., up to approximately half of the total residues of the protein), are particularly more localized on residues lining the cavities. Therefore, they are expected to have potential impact on protein function.

The evaluation of $\nabla \mathbf{V}_{ol}$ corresponding to a protein cavity requires algorithms for cavity volume calculations that results robust enough to structural distortions due to differential coordinate displacements. Our results were obtained using a combination of algorithms including convex hull, Delaunay triangulation and additional refinement steps (see section II.C). Other methods use predefined geometric shapes as spheres or prisms centered on user-specified coordinates. The number and size of these geometrical objects are subject to redefinitions and adjustments that can lead to cavity redefinitions even under small coordinate displacements. A convex hull drawn around user-specified atoms, instead of fixed coordinates, gives the required flexibility to keep the same definition of the cavity under differential structural distortions.

Figure 4 compares the robustness of different methods for the evaluation of $\frac{\partial \mathbf{V}_{ol}}{\partial \mathbf{Q}_i}$ (see 5) calculated by finite differences using different displacements. For the sake of simplicity, we exemplify our findings for Q_1, the lowest-frequency normal mode applied to HIVP-APO. Similar results are obtained for other coordinates and cases of study. We have considered the available software MDpocket[48] and Epock.[87] While MDpocket is based on Voronoi tessellation for cavity detection, Epock represents an implementation of grid-based methods where cavities are identified as free spaces accessible to a probe. Besides, we show results obtained using our procedure described in sections II.C and II.D but with Delaunay triangulation updated after structural distortion. MDpocket and Delaunay with retriangulation lead to rather unstable values of $\frac{\partial \mathbf{V}_{ol}}{\partial Q_1}$ with respect to changes in the relative displacements

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

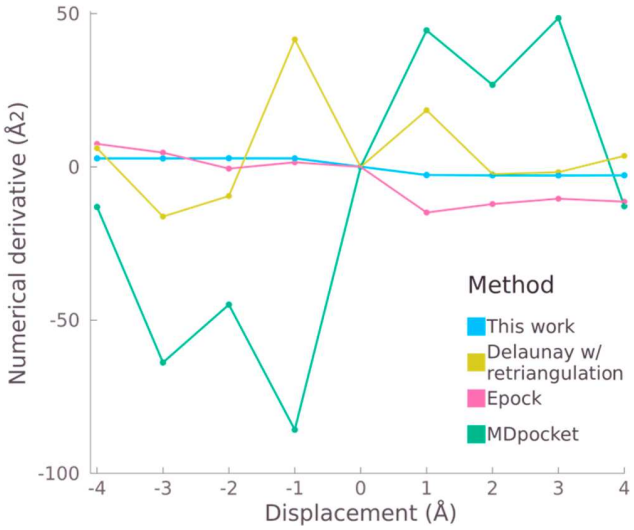


Figure 4. Comparison of robustness of different methods for the evaluation of $\frac{\partial \mathbf{V}_{ol}}{\partial Q_1}$ calculated by finite differences using different values of $X\Delta Q_1$.

used for their calculations. More stable results are achieved using Epock but only the procedure developed in this work results robust enough to achieve relative stable values of $\frac{\partial \mathbf{V}_{ol}}{\partial Q_1}$. A further comparison can be obtained by performing the overlap between $\nabla \mathbf{V}_{ol}$ vectors calculated using different ΔQ. Table I

Table I. Overlap Values between $\nabla \mathbf{V}_{ol}$ Vectors Obtained by Numerical Calculations Performed Using Different ΔQ Displacements and Volume Cavities Obtained Following the Procedure Described in This Work//Epock

	ΔQ			
ΔQ	2	4	6	8
2	1	0.80//0.70	0.82//0.50	0.73//0.43
4	0.80//0.70	1	0.90//0.75	0.81//0.64
6	0.82//0.50	0.90//0.75	1	0.88//0.81
8	0.73//0.43	0.81//0.64	0.88//0.81	1

compares results obtained using volume cavities calculated following the procedure described in this work and that of Epock. As it can be seen, larger values of overlaps are obtained between our $\nabla \mathbf{V}_{ol}$ vectors, indicating their robustness to changes in the relative displacements used for their calculations.

Cavity volumes are sensitive to relatively minor amino acids displacements, and accurately keeping track of these changes proved to be difficult with available software. Grid based methods proved to be inexact, and geometric methods, overly sensitive to slight variations. Therefore, we further test the consistency of our procedure for cavity volume calculations under small volume changes introduced by differential structural distortions. In order to do that, we consider $\nabla \mathbf{V}_{ol}$ expressed in Cartesian coordinates $\mathbf{q}$ applying the linear transformation between PCA modes $\mathbf{Q}$ and $\mathbf{q}$, given by eq 3, onto eq 4

$$\nabla \mathbf{V}_{ol} = \sum_{i=1}^{3N} \frac{\partial \mathbf{V}_{ol}}{\partial \mathbf{Q}_i} \mathbf{Q}_i = \sum_{i=1}^{3N} \frac{\partial \mathbf{V}_{ol}}{\partial \mathbf{Q}_i} \left(\sum_{j=1}^{N} \sum_{l=1}^{3} Q_{ji}^{l} \hat{\mathbf{q}}_j^{l} \right) = \sum_{j=1}^{N} \sum_{l=1}^{3} \frac{\partial \mathbf{V}_{ol}}{\partial q_j^{l}} \hat{\mathbf{q}}_j^{l} \quad (8)$$

where

$$Q_{ji}^{l} = \frac{\partial Q_i}{\partial q_j^{l}} \quad (9)$$

and

$$\frac{\partial \mathbf{V}_{ol}}{\partial q_j^{l}} = \sum_{i=1}^{3N} \frac{\partial \mathbf{V}_{ol}}{\partial \mathbf{Q}_i} \frac{\partial \mathbf{Q}_i}{\partial q_j^{l}} \quad (10)$$

The robustness of our algorithms for cavity volumes and numerical calculation of partial derivatives can be tested by comparing results obtained using eqs 8−10 with results obtained using direct finite diferences $\frac{\partial \mathbf{V}_{ol}}{\partial q_j^{l}} = \frac{\mathbf{V}_{ol} - \mathbf{V}_{ol,eq}}{\Delta q_j^{l}}$. In all cases, the overlap between $\nabla \mathbf{V}_{ol}$ calculated in these two ways was greater than 0.96. On the contrary, the corresponding overlap values using Epock for volume calculations were no larger than 0.57.

Figure 5 shows the relative residue displacements associated with the $\nabla \mathbf{V}_{ol}$ direction, calculated using eq 8. The most significant displacements correspond to residues localized on the surface of the cavities. While low-frequency collective fluctuations, involving the mobility of up to 50% of protein residues, contribute the most to volume cavity changes (see Figures 2 and 3), effective cancellations among the contributions of different modes to mobility of residues outside the cavities are observed. Therefore, changes in the cavity sizes do not actually involve visible large conformational changes throughout the protein structure but differential structural distortions and localized rearrangement of residues lining their surfaces (see the Supporting Information).

In the case of HIVP-APO, $\nabla \mathbf{V}_{ol}$ involves large displacements of residues belonging to the flap domain and located facing the cavity(∼43%), as well as residues belonging to the core domain (∼43%). Ligand binding slightly reduces the contributions of residues in the flap domain (∼38%). While $\nabla \mathbf{V}_{ol}$ of HIVP-HOLO is equally distributed between both monomers involving similar residues displacements on each of them, dynamics asymmetries are observed for HIVP-APO. This is due to asymmetric localization of the low-frequency modes that contribute the most to $\nabla \mathbf{V}_{ol}$.

EGFR active conformer presents a $\nabla \mathbf{V}_{ol}$ mainly localized on the β-strands and αC-helix of N lobe involving residues with side-chains pointing into the cavity (46%). It also involves relative motions of residues near the hinge region between N and C lobes (45%). The active EGFR conformation is relatively unstable. Previous studies[54] show that active−inactive conformational transition involves the opening of the N and C lobes to allow local unfolding at the hinge region, prior to the close of the lobes to restabilize in its inactive conformation. We observed that these motions do not significantly contribute to $\nabla \mathbf{V}_{ol}$. EGFR presents significant contributions of middle-range frequency modes to volume cavity changes, a feature that leads to a relative rigidity of its cavity.

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

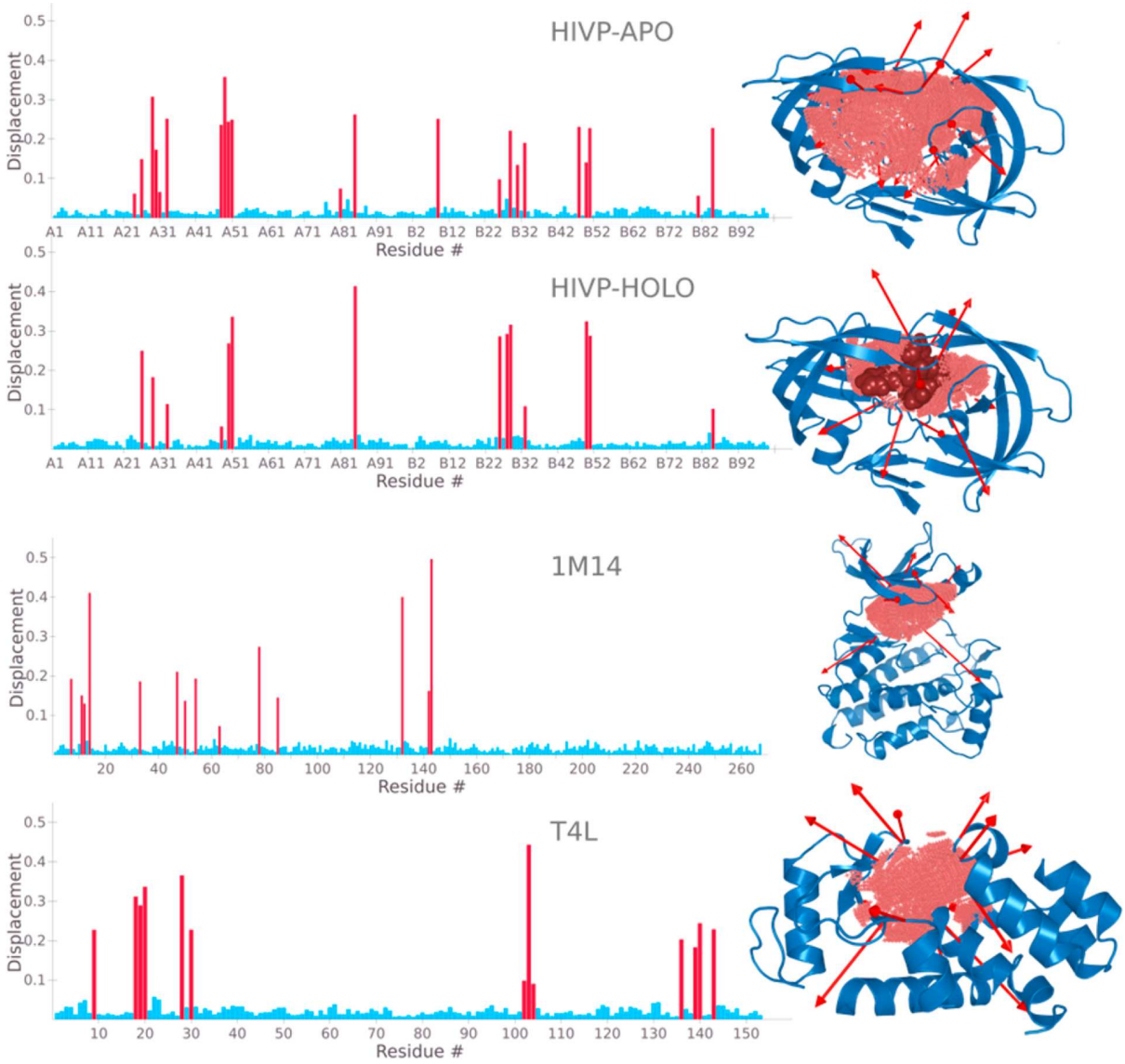


Figure 5. (left) Relative residue displacements associated with the $\nabla \mathbf{V}_{ol}$ direction obtained for HIVP-APO, HIVP-HOLO, EGFR, and T4L. Values of residues lining the cavities are colored in red and light blue is used for the rest of residues. (right) Corresponding structures represented by secondary structure motifs, with selected cavities colored in red, and porcupine plot of $\nabla \mathbf{V}_{ol}$ vector denoted with red arrows.

Low frequency modes involving N-terminal and C-terminal interdomain motions of hinge-bending, twisting, and shear represent significant contributions to $\nabla \mathbf{V}_{ol}$ in T4 lysozyme. This is in agreement with PCA studies that have previously reported[65,70] the large impact of these motions on the catalytic clef.

Within the frame of the quasi-harmonic analysis approximation,[88] the variation of the potential energy of a protein in the direction of $\nabla \mathbf{V}_{ol}$ can be used as a measure of flexibility of the cavity

$$\Delta E_{\nabla \mathbf{V}_{ol}} = \sum_{i=1}^{3N} \Delta E_{\mathbf{Q}_i} \tag{11}$$

being

$$\Delta E_{\mathbf{Q}_i} = \frac{1}{2} k_i c_i^2 \Delta X^2 \tag{12}$$

with $k_i = \frac{k_B T}{\lambda_i}$, where k_B is the Boltzmann constant and T is the absolute temperature (300 K). ΔX represents the displacement in the direction of $\nabla \mathbf{V}_{ol}$ relative the average structure used as reference in PCA calculations (see section II.B). Figure 6a displays $\Delta E_{\nabla \mathbf{V}_{ol}}$ as a function of the relative displacement ΔX for all cases of study. The order of flexibility of the cavities results HIVP-APO > T4L > EGFR > HIVP-HOLO. As it is expected, HIVP cavity becomes more rigid after ligand-binding. In order to give an insight into the values of relative flexibilities, Figure 6b shows the cumulative values of contributions of PCA modes (c_i^2) to $\nabla \mathbf{V}_{ol}$ (see eq 4) as a function of PCA mode frequencies. Proteins whose collective low frequency fluctuations participate the most on changes of cavity volume exhibit more flexible cavities. Despite the large contribution of these modes to displacements in the direction of $\nabla \mathbf{V}_{ol}$ (see Figure 2), they do not represent the main cost of energy in this direction. This can be seen in Figure 6c where we display the cumulative

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

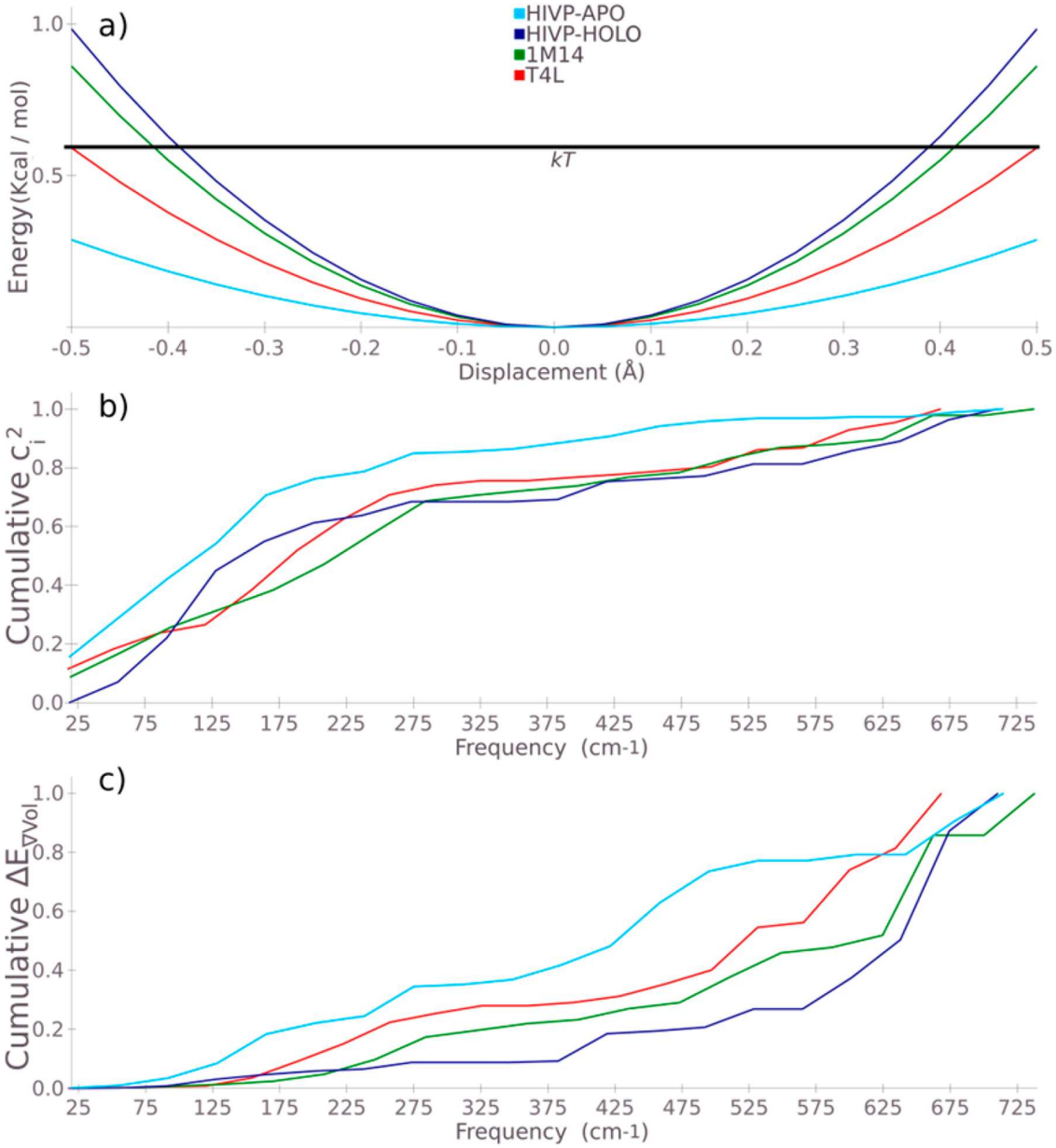


Figure 6. (a) Potential energy change in the direction of $\nabla \mathbf{V}_{ol}$; (b) cumulative values of contributions of PCA modes (c_i^2) to $\nabla \mathbf{V}_{ol}$ as a function of PCA mode frequencies; (c) cumulative values of $\frac{\Delta E_{\mathbf{Q}_i}}{\Delta E_{\nabla \mathbf{V}_{ol}}}$ as a function of PCA mode frequencies.

values of $\frac{\Delta E_{\mathbf{Q}_i}}{\Delta E_{\nabla \mathbf{V}_{ol}}}$ as a function of PCA mode frequencies. Low-frequency modes, that is, modes with frequencies up to 300 cm^{-1}, only contribute less than 40% to the total $\Delta E_{\nabla Vol}$ in the case of HIVP-APO, while their contributions are largely attenuated in the HIVP-HOLO by the presence of the ligand. Comparing our four cases of study, we can conclude that those relatively more flexible cavities (Figure 6a) are characterized by comparatively larger contributions of low frequencies modes to $\nabla \mathbf{V}_{ol}$ (Figure 6b) representing larger fraction of the total energy cost (Figure 6c).

IV. CONCLUSIONS

Protein fluctuations−cavity changes relationships have been explored analyzing cavity volumes gradients in terms of protein PCA modes. Our study provides complementary dynamics aspects to a direct evaluation of volume and protein fluctuations obtained during MD simulations. For this purpose, we make use of a combination of algorithms for cavity volume calculations robust enough for numerical differentiations.

In all our cases of study, the analysis of contributions of PCA modes to the volume gradient vector reveals major contributions of low frequency modes and minor contributions of high frequency modes. These low-frequency modes can involve fluctuations of up to approximately half of the total residues of the protein and 70% of residues of the cavities. Despite that, changes in the cavity sizes do not actually involve large conformational changes throughout the protein structure but localized rearrangement of residues lining their surfaces. This seems to take place due to effective cancellations among the contributions of different modes to mobility of residues outside the cavities.

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Considering the variation of the potential energy of a protein in the direction of $\nabla \mathbf{V}_{ol}$ as a measure of flexibility of the cavity, we observe that proteins whose collective low frequency fluctuations participate the most on changes of cavity volume exhibit more flexible cavities. Nevertheless, while low-frequency modes represent the main contributions to $\nabla \mathbf{V}_{ol}$, they do not represent the main cost of energy.

We consider these results as a first step to elucidate connections between protein fluctuations and volume cavity changes. The flexibility of protein cavities can impact on functional aspects like ligand affinities and binding promiscuities. Mutations that reduce the cost of energy in the direction of maxima changes in cavity volumes should increase cavity flexibility and, therefore, can potentially lead to increase substrate promiscuity.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00744.

Figure S1 and Table S1 (PDF)

Structural distortion of apo HIV protease in the direction of $\nabla \mathbf{V}_{ol}$ (movie_1hvr_apo.gif). Structural distortion of holo HIV protease in the direction of $\nabla \mathbf{V}_{ol}$ (movie_1hvr_holo.gif). Structural distortion of EGFR active conformer in the direction of $\nabla \mathbf{V}_{ol}$ (movie_1m14.gif). Structural distortion of T4 lysozyme in the direction of $\nabla \mathbf{V}_{ol}$ (movie_256l.gif). (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sfalberti@gmail.com (S.F.-A.).

ORCID

German P. Barletta: 0000-0002-0806-0812

Sebastian Fernandez-Alberti: 0000-0002-0916-5069

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem. Rev.* **2013**, *113* (8), 5871−5923.

(2) DesJarlais, R.; Sheridan, R.; Seibel, G.; Dixon, S.; Kuntz, I.; Venkataraghavan, R. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31* (4), 722−729.

(3) Chen, R.; Weng, Z. A Novel Shape Complementarity Scoring Function for Proteinprotein Docking. *Proteins: Struct., Funct., Genet.* **2003**, *51* (3), 397−408.

(4) Venkatachalam, C.; Jiang, X.; Oldfield, T.; Waldman, M. Ligandfit: A Novel Method for the Shape-Directed Rapid Docking of Ligands to Protein Active Sites.,. *J. Mol. Graphics Modell.* **2003**, *21* (4), 289−307.

(5) Hubbard, J.; Argos, P. Cavities and Packing at Protein Interfaces. *Protein Sci.* **1994**, *3* (12), 2194−2206.

(6) Sonavane, S.; Chakrabarti, P. Cavities and Atomic Packing in Protein Structures and Interfaces. *PLoS Comput. Biol.* **2008**, *4* (9), e1000188−09.

(7) Eyrisch, S.; Helms, V. What Induces Pocket Openings on Protein Surface Patches Involved in Protein−protein Interactions? *J. Comput.-Aided Mol. Des.* **2009**, *23* (2), 73−86.

(8) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8* (10), 23−30.

(9) Zhou, H.; Wlodek, S.; McCammon, J. Conformation Gating as a Mechanism for Enzyme Specificity. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (16), 9280−9283.

(10) Chothia, C.; Janin, J. Principles of Protein-Protein Recognition. *Nature* **1975**, *256*, 705−708.

(11) Argos, P. An Investigation of Protein Subunit and Domain Interfaces. *Protein Eng., Des. Sel.* **1988**, *2*, 101−113.

(12) Janin, J.; Miller, S.; Chothia, C. Surface, Subunit Interfaces and Interior of Oligomeric Proteins. *J. Mol. Biol.* **1988**, *204*, 155−164.

(13) Andersson, C.; Chen, B.; Linusson, A. Mapping of Ligand-Binding Cavities in Proteins. *Proteins: Struct., Funct., Genet.* **2009**, *78* (6), 1408−1422.

(14) Paramo, T.; East, A.; Garzón, D.; Ulmschneider, M. B.; Bond, P. J. Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: Trj-Cavity. *J. Chem. Theory Comput.* **2014**, *10* (5), 2151−2164.

(15) Seeliger, D.; Haas, J.; de Groot, B. Geometry-Based Sampling of Conformational Transitions in Proteins. *Structure* **2007**, *15*, 1482−1492.

(16) Kokh, D.; Czodrowski, P.; Rippmann, F.; Wade, R. Perturbation Approaches for Exploring Protein Binding Site Flexibility to Predict Transient Binding Pockets. *J. Chem. Theory Comput.* **2016**, *12*, 4100−4113.

(17) Stank, A.; Kokh, D. B.; Horn, M.; Sizikova, E.; Neil, R.; Panecka, J.; Richter, S.; Wade, R. C. TRAPP Webserver: Predicting Protein Binding Site Flexibility and Detecting Transient Binding Pockets. *Nucleic Acids Res.* **2017**, *45* (W1), W325−W330.

(18) Gunasekaran, K.; Ma, B.; Nussinov, R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins: Struct., Funct., Genet.* **2004**, *57* (3), 433−443.

(19) Pravda, L.; Berka, K.; Svobodová Vařeková, R.; Sehnal, D.; Banáš, P.; Laskowski, R. A.; Koča, J.; Otyepka, M. Anatomy of Enzyme Channels. *BMC Bioinf.* **2014**, *15* (1), 379.

(20) Pabon, N.; Camacho, C. Probing Protein Flexibility Reveals a Mechanism for Selective Promiscuity. *eLife* **2017**, DOI: 10.7554/eLife.22889.

(21) Barril, X.; Fradera, X. Incorporating Protein Flexibility into Docking and Structure-Based Drug Design. *Expert Opin. Drug Discovery* **2006**, *1* (4), 335−349.

(22) Monzon, A. M.; Zea, D. J.; Fornasari, M. S.; Fernandez-alberti, S.; Tosatto, S. C. E.; Parisi, G.; Saldano, T. E. Conformational Diversity Analysis Reveals Three Functional Mechanisms in Proteins. *PLoS Comput. Biol.* **2017**, *13*, e1005398.

(23) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press, 1987.

(24) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6679−6685.

(25) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(26) Brooks, C., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons Ltd, 1987.

(27) Ichiye, T.; Karplus, M. Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 205−217.

(28) Horiuchi, T.; Go, N. Projection of Monte Carlo and Molecular Dynamics Trajectories onto the Normal Mode Axes: Human Lysozyme. *Proteins: Struct., Funct., Genet.* **1991**, *10* (2), 106−116.

(29) van Aalten, D. M. F.; de Groot, B. L.; Findlay, J. B. C.; Berendsen, H. J. C.; Amadei, A. A Comparison of Techniques for Calculating Protein Essential Dynamics. *J. Comput. Chem.* **1997**, *18* (2), 169−181.

(30) Jolliffe, I. T. *Principal Component Analysis*, Second ed.; Springer-Verlag, 2002.

(31) David, C. C.; Jacobs, D. J. *Protein Dynamics*; Livesay, D. R., Ed.; Methods Mol. Biol.; Humana Press: Totowa, NJ, 2014; *Vol. 1084*.

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(32) Desdouits, N.; Nilges, M.; Blondel, A. Principal Component Analysis Reveals Correlation of Cavities Evolution and Functional Motions in Proteins. *J. Mol. Graphics Modell.* **2015**, *55*, 13−24.
(33) Connolly, M. Analytical Molecular Surface Calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.
(34) Levitt, D.; Banaszak, L. POCKET: A Computer Graphics Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino Acids. *J. Mol. Graphics* **1992**, *10*, 229−234.
(35) Kleywegt, G.; Jones, T. Detection, Delineation, Measurement and Display of Cavities in Macromolecular Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 178−185.
(36) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: Automatic and Efficient Detection of Potential Small Molecule-Binding Sites in Proteins. *J. Mol. Graphics Modell.* **1997**, *15*, 359−363.
(37) Huang, B.; Schroder, M. LIGSITEcsc: Predicting Ligand Binding Sites Using the Connolly Surface and Degree of Conservation. *BMC Struct. Biol.* **2006**, *6*, 19−29.
(38) Weisel, M.; Proschak, E.; Schneider, G. PocketPicker: Analysis of Ligand Binding-Sites with Shape Descriptors. *Chem. Cent. J.* **2007**, *1*, 7−23.
(39) Voss, N.; Gerstein, M. 3V: Cavity, Channel and Cleft Volume Calculator and Extractor. *Nucleic Acids Res.* **2010**, *38*, W555−W562.
(40) Chwastyk, M.; Jaskolski, M.; Cieplak, M. The Volume of Cavities in Proteins and Virus Capsids. *Proteins: Struct., Funct., Genet.* **2016**, *84* (9), 1275−1286.
(41) Laskowski, R. SURFNET: A Program for Visualizing Molecular Surfaces, Cavities, and Intermolecular Interactions. *J. Mol. Graphics* **1995**, *13*, 323−330.
(42) Brady, G.; Stouten, P. Fast Prediction and Visualization of Protein Binding Pockets with PASS. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 383−401.
(43) Kawabata, T.; Go, N. Detection of Pockets on Protein Surfaces Using Small and Large Probe Spheres to Find Putative Ligand Binding Sites. *Proteins: Struct., Funct., Genet.* **2007**, *68*, 516−529.
(44) Masood, T.; Sandhya, S.; Chandra, N.; Natarajan, V. CHEXVIS: A Tool for Molecular Channel Extraction and Visualization. *BMC Bioinf.* **2015**, *16* (1), 1−19.
(45) Petřek, M.; Košinová, P.; Koča, J.; Otyepka, M. MOLE: A Voronoi Diagram-Based Explorer of Molecular Channels, Pores, and Tunnels. *Structure* **2007**, *15* (11), 1357−1363.
(46) Pavelka, A.; Sebestova, E.; Kozlikova, B.; Brezovsky, J.; Sochor, J.; Damborsky, J. CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2016**, *13* (3), 505−517.
(47) Yaffe, E.; Fishelovitch, D.; Wolfson, H.; Halperin, D.; Nussinov, R. MolAxis: Efficient and Accurate Identification of Channels in Macromolecules. *Proteins: Struct., Funct., Genet.* **2008**, *73* (1), 72−86.
(48) Schmidtke, P.; Bidon-Chanal, A.; Luque, F.; Barril, X. MDpocket: Open-Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011**, *27* (23), 3276−3285.
(49) Berka, K.; Hanák, O.; Sehnal, D.; Banáš, P.; Navrátilová, V.; Jaiswal, D.; Ionescu, C. M.; Svobodová Vařeková, R.; Koča, J.; Otyepka, M. MOLEonline 2.0: Interactive Web-Based Analysis of Biomacromolecular Channels. *Nucleic Acids Res.* **2012**, *40* (W1), 222−227.
(50) Chang, C.-E.; Shen, T.; Trylska, J.; Tozzini, V.; McCammon, J. A. Gated Binding of Ligands to HIV-1 Protease: Brownian Dynamics Simulations in a Coarse-Grained Model. *Biophys. J.* **2006**, *90* (11), 3880−3885.
(51) Gustchina, A.; Weber, I. T. Comparison of Inhibitor Binding in HIV-1 Protease and in Non-Viral Aspartic Proteases: The Role of the Flap. *FEBS Lett.* **1990**, *269* (1), 269−272.
(52) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J.; Whittle, P.; Danley, D.; Geoghegan, K.; et al. X-Ray Analysis of HIV-1 Proteinase at 2.7 A Resolution Confirms Structural Homology among Retroviral Enzymes. *Nature* **1989**, *342*, 299−302.
(53) Wlodawer, A.; Erickson, J. W. Structure Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543−585.
(54) Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. Transitions to Catalytically Inactive Conformations in EGFR Kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (18), 7270−7275.
(55) Ferguson, K. Active and Inactive Conformations of the Epidermal Growth Factor Receptor. *Biochem. Soc. Trans.* **2004**, *32* (5), 742−745.
(56) Kornev, A.; Taylor, S. Defining the Conserved Internal Architecture of a Protein Kinase. *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804* (3), 440−444.
(57) Kalia, M. Biomarkers for Personalized Oncology: Recent Advances and Future Challenges. *Metab., Clin. Exp.* **2015**, *64* (3), S16−S21.
(58) Arteaga, C.; Engelman, J. ERBB Receptors: From Oncogene Discovery to Basic Science to Mechanism-Based Cancer Therapeutics. *Cancer Cell* **2014**, *25* (3), 282−303.
(59) Fabbro, D. 25 Years of Small Molecular Weight Kinase Inhibitors: Potentials and Limitations. *Mol. Pharmacol.* **2015**, *87* (5), 766−775.
(60) Faber, H.; Matthews, B. A Mutant T4 Lysozyme Displays Five Different Crystal Conformations. *Nature* **1990**, *348*, 263−266.
(61) Matthews, B.; Remington, S. The Three Dimensional Structure of the Lysozyme from Bacteriophage T4. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 4178−4182.
(62) Dixon, M.; Nicholson, H.; Shewchuk, L.; Baase, W.; Matthews, B. Structure of a Hingebending Bacteriophage T4 Lysozyme Mutant, Ile3RPro. *J. Mol. Biol.* **1992**, *227*, 917−933.
(63) Zhang, X.; Wozniak, J.; Matthews, B. Protein Flexibility and Adaptability Seen in 25 Crystal Forms of T4 Lysozyme. *J. Mol. Biol.* **1995**, *250*, 527−552.
(64) Mchaourab, H.; Oh, K.; Fang, C.; Hubbell, W. Conformation of T4 Lysozyme in Solution. Hinge-Bending Motion and the Substrate-Induced Conformational Transition Studied by Site-Directed Spin Labeling. *Biochemistry* **1997**, *36*, 307−316.
(65) McCammon, J.; Gelin, B.; Karplus, M.; Wolynes, P. The Hinge Bending Mode in Lysozyme. *Nature* **1976**, *262*, 325−326.
(66) Brooks, B.; Karplus, M. Normal Modes for Specific Motions of Macromolecules: Application to the Hinge-Bending Mode of Lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 4995−4999.
(67) Hayward, S.; Kitao, A.; Berendsen, H. Model Free Methods of Analyzing Domain Motions in Proteins from Simulation: A Comparison of Normal Mode Analysis and Molecular Dynamics Simulation of Lysozyme. *Proteins: Struct., Funct., Genet.* **1997**, *27*, 425−437.
(68) Hayward, S.; Berendsen, H. Systematic Analysis of Domain Motions in Proteins from Conformational Change: New Results on Citrate Synthase and T4 Lysozyme. *Proteins: Struct., Funct., Genet.* **1998**, *30*, 144−154.
(69) de Groot, B.; Hayward, S.; van Aalten, D.; Amadei, A.; Berendsen, H. Domain Motions in Bacteriophage T4 Lysozyme; a Comparison between Molecular Dynamics and Crystallographic Data. *Proteins: Struct., Funct., Genet.* **1998**, *31*, 116−127.
(70) Hub, J. S.; de Groot, B. L. Detection of Functional Modes in Protein Dynamics. *PLoS Comput. Biol.* **2009**, *5* (8), e1000480.
(71) Salomon-Ferrer, R.; Case, D.; Walker, R. An Overview of the Amber Biomolecular Simulation Package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198−210.
(72) Case, D.; Cheatham, T. I.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(73) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(74) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(75) Ponder, J.; Case, D. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27−85.

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(76) Maier, J.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(77) Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.
(78) Mukherjee, G.; Patra, N.; Barua, P.; Jayaram, B. A Fast Empirical GAFF Compatible Partial Atomic Charge Assignment Scheme for Modeling Interactions of Small Molecules with Biomolecular Targets. *J. Comput. Chem.* **2011**, *32* (5), 893−907.
(79) Jakalian, A.; Jack, D.; Bayly, C. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23* (16), 1623−1641.
(80) Van Aalten, D.; De Groot, B.; Findlay, J.; Berendsen, H.; Amadei, A. A Comparison of Techniques for Calculating Protein Essential Dynamics. *J. Comput. Chem.* **1997**, *18*, 169−181.
(81) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17* (4), 412−425.
(82) Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H. The Quickhull Algorithm for Convex Hulls. *ACM Trans. Math. Softw.* **1996**, *22* (4), 469−483.
(83) Simões, T.; Lopes, D.; Dias, S.; Fernandes, F.; Pereira, J.; Jorge, J.; Bajaj, C.; Gomes, A. Geometric Detection Algorithms for Cavities on Protein Surfaces in Molecular Graphics: A Survey. *Comput. Graph. Forum* **2017**, *36*, 643.
(84) Krone, M.; Kozlíková, B.; Lindow, N.; Baaden, M.; Baum, D.; Parulek, J.; Hege, H. C.; Viola, I. Visual Analysis of Biomolecular Cavities: State of the Art. *Comput. Graph. Forum* **2016**, *35* (3), 527−551.
(85) Bhattacharyya, A. On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distribution. *Bull. Calcutta Math. Soc.* **1943**, *35*, 99−110.
(86) Aherne, F.; Thacker, N.; Rockett, P. The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data. *Kybernetika* **1998**, *33* (4), 363−368.
(87) Laurent, B.; Chavent, M.; Cragnolini, T.; Dahl, A.; Pasquali, S.; Derreumaux, P.; Baaden, M.; Sansaom, M. S. P. Epock: Rapid Analysis of Protein Pocket Dynamics. *Bioinformatics* **2015**, *31* (9), 1478−1480.
(88) Brooks, B.; Janezic, D.; Karplus, M. Harmonic Analysis of Large Systems. I. Methodology. *J. Comput. Chem.* **1995**, *16* (12), 1522−1542.

DOI: 10.1021/acs.jctc.7b00744
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX