Vibrational Energy in Proteins Correlates with Topology

L. Maggi*†, P. Carloni†‡§||, G. Rossetti*†¶⊥

† Computational Biomedicine Section, Institute of Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

‡ Institute for Neuroscience and Medicine INM-11, Forschungszentrum Jülich, 52428 Jülich, Germany;

§ Department of Physics, RWTH Aachen University, 52078 Aachen, Germany

|| Department of Neurology, University Hospital Aachen, 52078 Aachen, Germany

¶ Division Computational Science − Simulation Laboratory Biology, Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

⊥ Department of Oncology, Hematology and Stem Cell Transplantation, University Hospital Aachen, RWTH Aachen University, 52074 Aachen, Germany

*l.maggi@fz-juelich.de,*g.rossetti@fz-juelich.de

ABSTRACT The exchange of vibrational energy in proteins is crucial for their function. Here, we establish a connection between quantities related to it with geometry-based properties such as the proteins' residues coordination number. This relation is proven by molecular simulation in a neuro-pharmacologically relevant transmembrane receptor. The connection demonstrated here paves the way to studies of protein allostery and conformational changes based solely on protein structure.
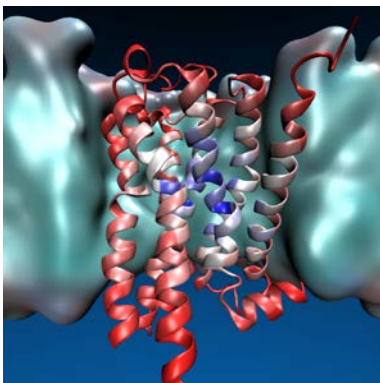
The vibrational energy exchange (*Evib_xc)* among proteins residues is an important probe for protein function[1]. Changes in *Evib_xc* are associated with allostery[2], conformational rearrangement[3] and with heat dissipation subsequent to a chemical reaction[4]. For instance, a recent experimental study on bovine serum albumin has shown that *Evib*_xc is modified upon allosteric ligand binding[5]. Furthermore, anti-Stokes resonance Raman spectroscopy revealed *Evib* relaxation subsequent to the photodissociation of carbon monoxide in myoglobin[4].

The complex protein topology and the broad range of different chemical interactions within bio-macromolecules makes these important processes highly anisotropic and challenging to investigate[6].

Molecular simulations[2, 7] and time-resolved spectroscopies may lead to detect *Evib_xc* in specific proteins' residues[4,8]. Both computational[9] and experimental[10] studies suggested a connection between *Evib_xc* and protein topology. In the following, we use molecular simulation to show unambiguously such connection, allowing to identify protein regions with low and high efficiency in *Evib_xc*. This in turn permits to interrogate protein topologies to obtain information on regions playing roles for biological functions, allosteric communication, or being important hubs for structural stability and/or conformational transitions[3, 5]. In order to study this relation, we have calculated properties related to *Evib_xc*, within the framework of the linear response Theory (See Section S1). We used for this equilibrium Molecular Dynamics simulation. This procedure differs from previous ones (Refs. 2-4), in which Non-equilibrium Molecular Dynamics simulations have been employed.

Our demonstration is based on two basic, well tested approximations: (i) We assume the potential energy surface of the protein as parabolic via the Quasi Harmonic Approximation[11](QHA, See Section S2 for details). We adopt here a quantum description of the normal modes; (ii) We consider only the centers of mass for each residue. The choice of including the mass center allows to decrease greatly the number of degrees of freedom. This greatly reduces the complexity of the system, albeit makes our description less detailed. In addition, it renders QHA a rather reliable methodology, as approximating the energy surface as harmonic is better in simple systems with few degrees of freedom than in complex ones. Finally, this choice allows to include (albeit indirectly) all the interactions coming from both the side-chains and the backbone, since the center of mass of a residue is affected by all of them. In conclusion, this choice allows us to take in to account indirectly all of the intramolecular interactions while keeping the system simple and amenable for the QHA. All the side-chains and main-chains interactions' effects are however included indirectly

Let us first introduce a quantity related to *Evib_*xc processes. This is the average time needed by each protein residue to exchange *Evib* with its neighbors. We call this the Characteristic Diffusion Time (CDT). For a specific residue *i*, CDT is defined as:

$$CDT_i = \frac{\sum_{j=1}^{N} \tau^{ij}}{N}$$

where j runs along the *N* protein residues within a cutoff from residue *i*, and τ^{ij} is the average time for the vibrational energy transfer between residue *i* and any other protein residue *j,* as calculated in[2] (See Section S1 for details). CDT*$_i$* is obviously related to *Evib_xc:* it is short when the *Evib_xc* is fast and efficient and large in the opposite case.

The cutoff is here set to 1.2 nm. We have chosen this value because this is the cut-off of real-space electrostatic and van der Waals interactions in our simulations. Beyond these distances, these contributions are not calculated. Hence, setting the cutoff at 1.2 nm allows to consider entirely these contributions to the non-bonded interactions.

Next, we show that CDT_i anti-correlates with the mean square displacement of the center of mass of each residue i (MSD_i). The latter indeed depends on the number of interacting residues, namely the protein topology, and on the strength (i.e. the chemical nature) of each interaction.

To show that the CDT_i values are non-linearly anti-correlated to the set of MSD_i ones, we first link the CDT_i value for residue i to the E_i, the vibrational energy for the same residue:

$$E_i = \sum_l \left| \epsilon_l^i \right|^2 E(\omega_l)$$

where the sum over the l=1,2,... normal modes is calculated within the QHA. $E(\omega_l)$ is the energy associated with the l-th normal mode. $|\epsilon_l^i|^2$ is the sum of components of the l-th eigenvector related to the degrees of freedom of the i-th residue. $|\epsilon_l^i|^2$ is a property of residue i and it tells how much the l-th normal mode contributes to E_i.

The *Evib*_xc process for residue i involves a transition between harmonic oscillator states. The smaller the energy gap between these states ($\hbar\omega$), the more probable such transition is and, consequently, the easier the process is. Thus, the larger is $|\epsilon_l^i|^2$ in the low-frequency modes, the faster is the energy exchange for residue i and, consequently, the smaller CDT_i.

Similarly to E_i, the MSD_i for residue i can be written as a linear combination of the protein's normal modes mean square displacement (MSD_l), with the $|\epsilon_l^i|^2$ values as coefficients:

$$MSD_i = \sum_l |\epsilon_l^i|^2 \, MSD_l$$

MSD_l is inversely proportional to the frequency of each normal mode. Hence, the larger the MSD_i, the larger $|\epsilon_i^i|^2$ on the low-frequency modes, the easier the *Evib* exchange process and the shorter CDT_i for this residue. Conversely, the shorter MSD_i, the larger the CDT_i. This demonstrates the relation between the two quantities.

While MSD_i is related to structural and dynamical features, it is not a quantity that is easy to relate to protein topology. Hence, we introduce here a new quantity, the number of interacting residues with residue *i* (N_i) within our chosen cutoff. Under the assumption that the strength of all neighbor residues is the same, this assumption is validated numerically for other three proteins then the one presented in the following (See Supporting information, Section S3 and Fig S1), N_i is anti-correlated with MSD_i: the latter increases upon a decrease of N_i. This establishes a relation between the *Evib_xc* and the protein topology via the intuitive quantity N_i. We show next that this correlation exists in a one of the proteins investigated here, the human Muscarinic receptor M2. Our analysis allows also to relate protein regions with distinct *Evib* exchange to receptor's activation and function.

M2 belongs to class A G-protein Coupled Receptors (GPCRs). GPCRs can assume an active and an inactive conformational states and trigger a signal cascade within the cell by binding different macromolecular partners[12]. M2 consists of 7 transmembrane helices (H 1-7), connected by intracellular and extracellular loops, along with an intracellular helix (H8). Upon binding to its agonists, M2 changes its conformation from inactive state to active state[13]. In the latter, the alpha-

helix 6 facing the cytoplasm side moves outward away from the center of the helix bundle[1] (Fig. 1A).
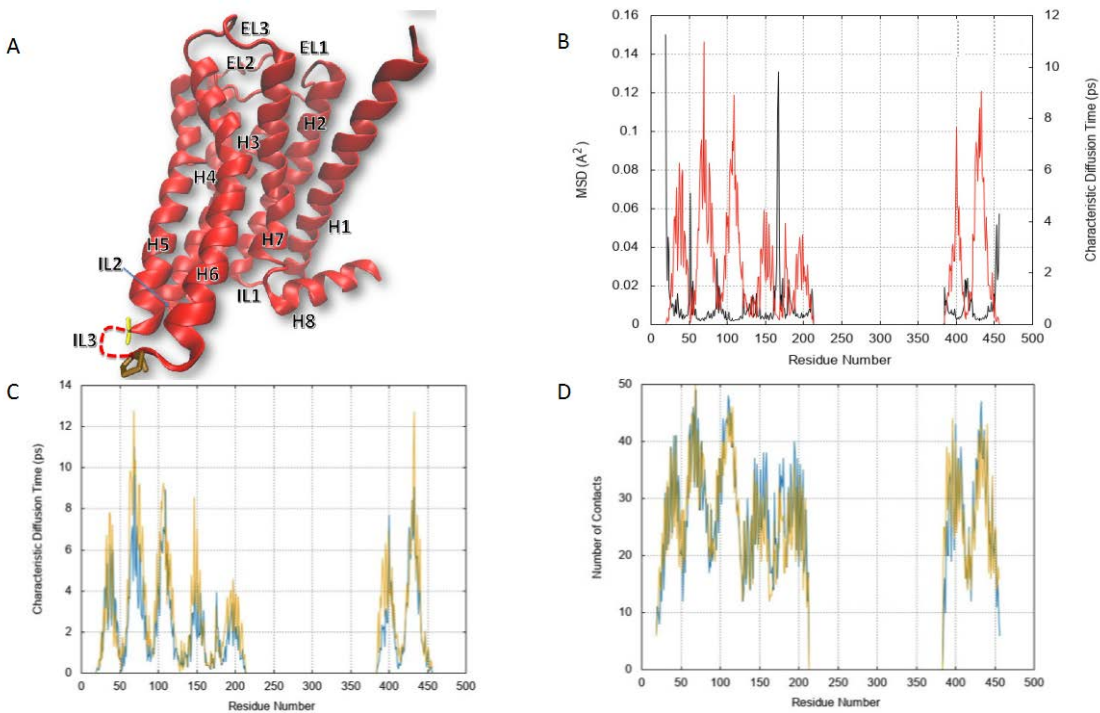

Figure 1: M2 Muscarinic receptor in the active state. The trans-membrane helices are labeled with "H", the intracellular and extracellular loops with "IL" and "EL" respectively (A). CDT_i (red line) and MSD_i (black line) for M2 active state along the residue numbers (B). CDT_i for M2 active state (blue) and M2 inactive state (yellow) plotted along the residue numbers (C). Number of Contacts for M2 active state (blue) and M2 inactive state (yellow) plotted along the residue

---

[1] Similarly, large rearrangements are observed not only for M2, but also for other class A GPCR[15].

numbers (D). The chosen cut-off for determining the number interacting residues has been set to 1.2 nm. The gap between residue 212 and residue 383 is due to the intracellular loop (IL3).

In our molecular dynamics (MD) simulations, we consider the X-ray structure of the receptor in its inactive [14] (PDB ID: 3UON) and active [13] (PDB ID: 4MQS) states. The two structures are bound to ligands that stabilize the specific conformations (See Section S4 for details). The proteins are embedded in a membrane composed by different type of lipids whose concentrations mimic the ones of a prefrontal cortex cellular membrane (See Tab. S1). The environment in which the protein is embedded is highly anisotropic. However, here we consider only residue-residue interactions within the protein itself. This means the external environment has not been taken into account explicitly. In our approximated picture those interactions are included in narrow energy span which make our assumption reliable (See Section in S3 and Fig. S1).

A long intracellular loop connecting the alpha helix 5 and 6 can be considered a distinct domain with respect to the seven helices bundle. The structure for this loop was obtained by *ab initio* modeling. We should point out that the accuracy of this structural prediction is rather low, as the loop consists of more than 160 residues [15] (See Section S4 for details). Hence, QHA is rather unreliable for residues comprised in the loop. However, it has been proven for another GPCR (beta2-adrenergic receptor) that the loop does not influence the structural features[16]. Hence, the investigation of this loop was excluded from our analysis. Next, we perform 50 ns MD atomistic simulations at 300 K and 1 atm of pressure in explicit solvent. An independent 50-ns replica has been performed in order to validate further our findings (See Section S4 for details and Fig. S2). Convergence analysis is offered in Section S4 Fig. S3 and Tab. S2. Finally, we calculated CDT_i, MSD_i and N_i values for each protein residue i based on the QHA analysis of the MD trajectories.

A comment is in order regarding the use of the QHA. The potential energy of a protein (such as the M2 receptor here) is obviously not harmonic. Hence, it is impossible to describe the transition from inactive to active state of the M2 receptor with such potential energy. Here, we investigate separately two very important conformational states, the active one and the inactive one. In those states, the potential energy surface is located in basins[17]. The QHA can be considered a rough, yet reasonable, local approximation of these basins. Hence, the use of QHA in each of these states is justified.

Fig. 1B and Fig. S4 in Section S5 show the non-linear anticorrelation between CDT_i and the MSD_i for both states, confirming the validity of our reasoning. Fig. 2A-B shows the non-linear correlation of CDT_i with N_i and non-linear anti-correlation of MSD_i with N_i for both states. This supports our assumption that each neighbor residue exerts roughly the same interaction on residue i. We performed the same analysis on three different cyto-plasmatic proteins which have different topology than the M2 receptor. The same correlations between CDT, MSD and number of contacts were found also for these proteins (See Section S6 and Fig. S5-S7). It should be noticed the CDT calculation also the contacts between the protein and the ligand (as well as the one between the protein and the solvent, or the membrane) would have provided us with information about the energy exchange between the protein and other components of the system. However, the paper focuses on the vibrational energy exchange within the protein under the effect of the ligand, the solvent and the membrane. Hence, the effect of all the latter are considered indirectly, via QHA as pointed out in the Supporting Information, Section S2. Hence, these contacts were not included.
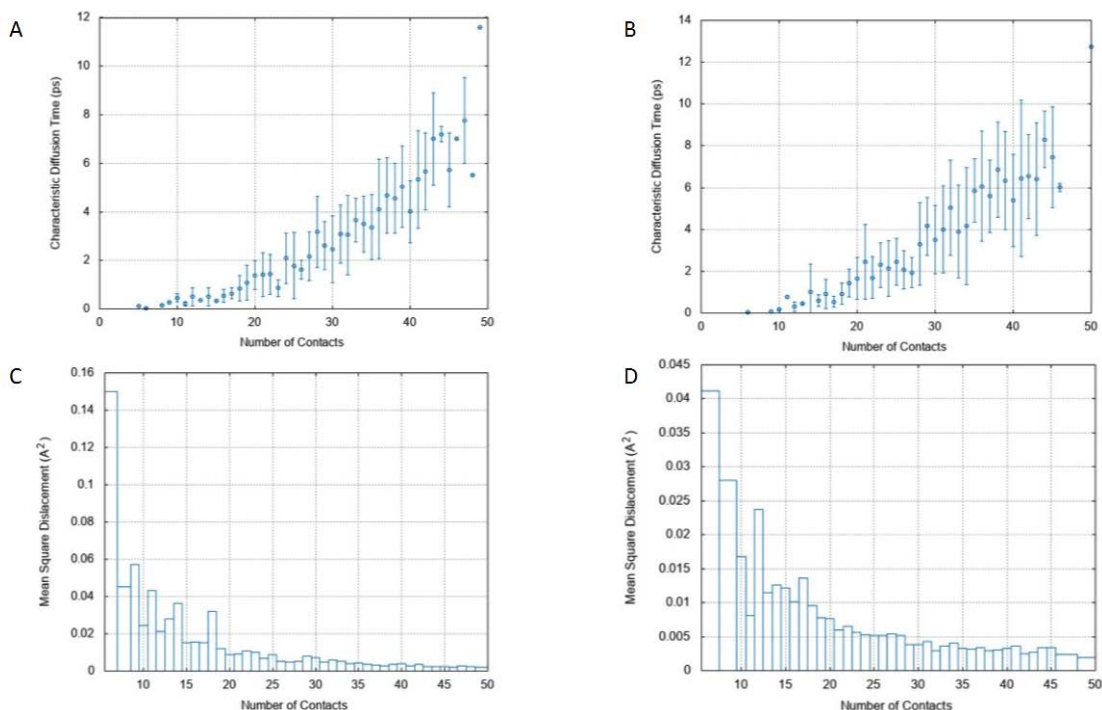
Figure 2: CDT_i against the number of contacts in the active (A) and the inactive state (B). The plots show the quantity averaged for each value of the number of contacts. The error bars point out a significant variability around the mean values. This is caused by the different nature of interactions. MSD_i is plotted against the number of contacts in the active (C) and the inactive (D) states. The plots show the quantity averaged for each value of number of contacts.

Common traits across the inactive and active states. The CDT_i values of the transmembrane region[2] are larger (ranging from 2.5 to 13 ps) than those of the rest of the protein (ranging from 0.5 ps to 2.5 ps) for both states (Fig. 1C). Therefore, this region exchanges vibrational energy

---

[2] Defined as the residue comprised within certain values of z_{max} and z_{min}. Changes in these definitions turned out not to affect the conclusions of this section (data not shown)

slower than the rest. Consistently, the values of $|\varepsilon_I^i|^2$ averaged over the number of residues ($<|\varepsilon_I^i|^2>$) related to these two regions show opposite trends. In the frequency region below 2KbT (with T=300K) $<|\varepsilon_I^i|^2>$ of the transmembrane region presents shorter value than those related to the rest of the protein. In contrast these trends are completely reversed in the frequency regime above 2KbT (Fig 3a-b).
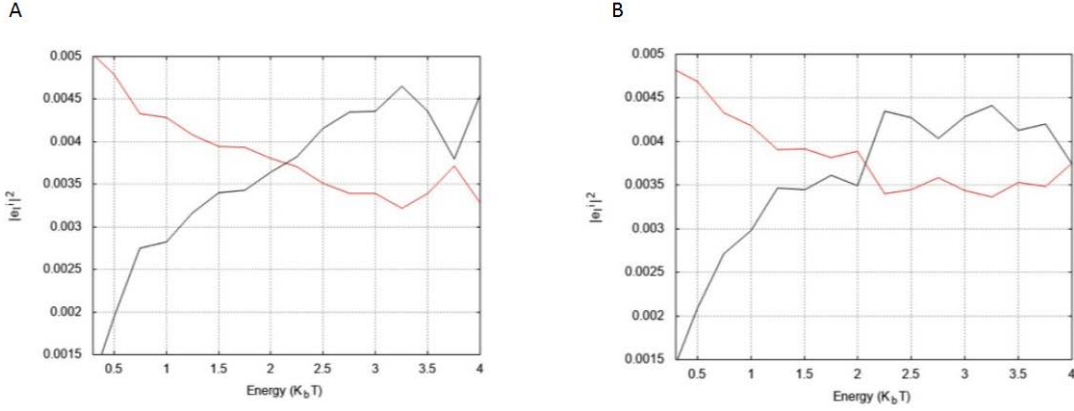
A   B

Figure 3: $<|\varepsilon_I^i|^2>$ along frequencies, measured in KbT unit, related to different region of the protein. Transmembrane region (black line), non-transmembrane regions (red line) for M2 active (A) and inactive (B) states

   This means that the energy exchange of residues in the transmembrane region is slowed down, significantly hampering all the possible conformational changes. Although the latter take place on a long-time scale, they usually need multiple elementary shorter-time scale steps, which involve the energy exchange between the bio-molecule degrees of freedom [18].

   Furthermore, in both states, the CDT values of the transmembrane helices exhibit a gaussian-like profile, whose maximum is roughly localized in their middle (see Fig. 1A and 3A). Namely, the maxima are localized onto highly to fairly conserved residues ($W^{6.48}, S^{7.46.}, I^{1.48}, A^{3.38}, W^{4.50}$,

$D^{2.50}$, $P^{5.50}$).[3] Except for $P^{5.50}$, all of them are clustered within 1.2 nm from the absolute maximum ($D^{2.50}$, see Fig. 4). All these residues play a fundamental role for the stability and the function of class A GPCRs [19]. In particular, single point mutation experiment has shown mutating $D^{2.50}$ in A disrupt completely the basal activity of Class A GPCR A_2 Adenosine Receptor [20].
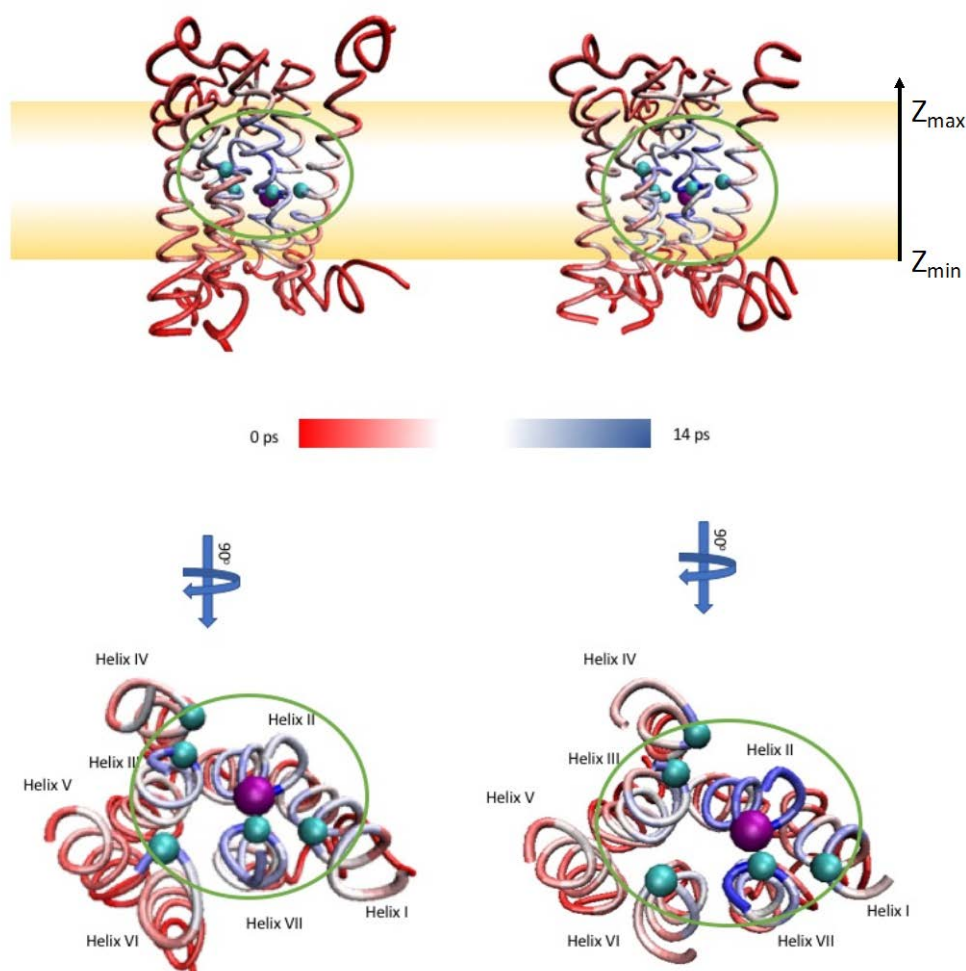

Figure 4: CDT$_i$ values for M2 active state (right) and the M2 inactive state projected onto the 3D structures. The shaded orange colored strip has been drawn to schematically show where the membrane is located. The molecules inside the receptors are also shown, in green.

---

[3] We use here a Ballesteros-Weinstein numbering[19].

In the light of our findings, these experimental results could be interpreted as a change of inter-residues interactions due to the mutation. By modifying the shape and the chemical nature of the residues, also its packing and the energy exchange in the transmembrane region might be influenced, therefore both affecting its ability to transduce the signal and compromising the entire receptor function. This analysis does not show any specific residue belonging to the extracellular and the intracellular compartments that have features differing from the general picture previously exposed.

Differences. While the trend of the CDT_i values between the transmembrane region and the rest of the protein are the same in the two states, their absolute values are different (Fig. 1C). The inactive state ones are larger and hence the *Evib_xc* processes are slower. This might be associated with the fact that the inactive state features larger thermodynamic stability than the active one [21]. This larger CDT_i values in the inactive state can be ascribed to larger vibrational mode frequencies in such state (Fig. S2). This determines larger energy gaps between the harmonic oscillators, which in turns hamper *Evib_xc* processes. The increase in vibrational mode frequencies in the inactive state are caused by the strengthening, on average, of the inter-residue interactions, upon conformational transition. As result the whole structure becomes stiffer, i.e. characterized by increased vibrational frequency (See Section S7 Fig. S8). These results are consistent with experimental studies, on small helical peptides, which show that energy transport is less efficient in stiffer structures[22]. We finally notice that calculating the average CDT_i for each of the seven transmemebrane helices (See Section S8 Tab. S3), The largest decrease among the average CDT_i occurs at the region of the protein known to experience the largest rearrangements upon activation, which is the bottom of the alpha helix 6: this value is doubled on passing from the inactive state (1.7 ps) to the active one (3.5 ps) and hence N_i decreases (from 30 and 22).

In conclusion, we have shown a general qualitative connection between the *Evib* exchange and proteins' topology. Residues having a short number of contacts, presents a higher efficiency to transfer *Evib* than others, because the contribution to their energy comes mostly from low-frequency modes.

Experimental results [10] have shown that the residues featuring a larger Surface Accessible Solvent Area (SASA) present an efficient *Evib_xc*. SASA clearly anti-correlates with the number of contacts[23]. Hence this establishes a connection between our findings and the experimental ones. Showing our analysis can explain the results of ref. [23] from a theoretical point of view

Our prediction has been shown to be valid place a test-case system, the Muscarinic M2 receptor in its inactive and active states. As a bonus, this analysis has shown that specific energy exchange properties and topological features may play an important role for the stability and the activation process of the receptor. This work may help to investigate protein allostery and protein rearrangements using structural information.

AUTHOR INFORMATION:

Corresponding Authors

*E-mail: l.maggi@fz-juelich.de,
*E-mail:g.rossetti@fz-juelich.de

Notes

The authors declare no competing financial interests.

Supporting Information Available: Simulation details, Convergence Analysis, Quasi harmonic approximation, Characteristic Diffusion Time, Analysis of protein with different topology, Residue Interactions Approximation, Mean Square Displacement, RMSF and Characteristic Diffusion Time in M2 (in)active state, Normal Mode Frequencies, Helix averaged CDT and Membrane Lipid Composition

REFERENCES:

(1) Leitner, D. M.; Straub, J. E. *Proteins: energy, heat and signal flow*. CRC Press: 2009.
(2) Buchenberg, S.; Leitner, D. M.; Stock, G. Scaling Rules for Vibrational Energy Transport in Globular Proteins. *J Phys Chem Lett* **2016,** *7*, 25-30.
(3) Bastida, A.; Zuniga, J.; Requena, A.; Miguel, B.; Candela, M. E.; Soler, M. A. Conformational Changes of Trialanine in Water Induced by Vibrational Relaxation of the Amide I Mode. *J Phys Chem B* **2016,** *120*, 348-357.
(4) Mizutani, Y.; Kitagawa, T. Direct observation of cooling of heme upon photodissociation of carbonmonoxy myoglobin. *Science* **1997,** *278* (5337), 443-446.
(5) Li, G. F.; Magana, D.; Dyer, R. B. Anisotropic energy flow and allosteric ligand binding in albumin. *Nat Commun* **2014,** *5*.

(6) Leitner, D. M. Energy flow in proteins. *Annu Rev Phys Chem* **2008,** *59*, 233-259.

(7) Leitner, D. M.; Buchenberg, S.; Brettel, P.; Stock, G. Vibrational energy flow in the villin headpiece subdomain: Master equation simulations. *J Chem Phys* **2015,** *142*.

(8) Fujii, N.; Mizuno, M.; Ishikawa, H.; Mizutani, Y. Observing Vibrational Energy Flow in a Protein with the Spatial Resolution of a Single Amino Acid Residue. *J Phys Chem Lett* **2014,** *5*, 3269-3273.

(9) Mino-Galaz, G. A. Allosteric Communication Pathways and Thermal Rectification in PDZ-2 Protein: A Computational Study. *J Phys Chem B* **2015,** *119*, 6179-6189.

(10) Kondoh, M.; Mizuno, M.; Mizutani, Y. Importance of Atomic Contacts in Vibrational Energy Flow in Proteins. *J Phys Chem Lett* **2016,** *7*, 1950-1954.

(11) Bialek, W.; Goldstein, R. F. Do Vibrational Spectroscopies Uniquely Describe Protein Dynamics - the Case for Myoglobin. *Biophys J* **1985,** *48*, 1027-1044.

(12) Hanlon, C. D.; Andrew, D. J. Outside-in signaling - a brief review of GPCR signaling with a focus on the Drosophila GPCR family. *J Cell Sci* **2015,** *128*, 3533-3542.

(13) Kruse, A. C.; Ring, A. M.; Manglik, A.; Hu, J. X.; Hu, K.; Eitel, K.; Hubner, H.; Pardon, E.; Valant, C.; Sexton, P. M.; Christopoulos, A.; Felder, C. C.; Gmeiner, P.; Steyaert, J.; Weis, W. I.; Garcia, K. C.; Wess, J.; Kobilka, B. K. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **2013,** *504*, 101.

(14) Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; Kobayashi, T. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **2012,** *482*, 547-U147.

(15) Lee, J.; Freddolino, P. L.; Zhang, Y. Ab initio protein structure prediction. In *From protein structure to function with bioinformatics*, Springer: 2017; pp 3-35.

(16) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Yao, X. J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K. GPCR engineering yields high-resolution structural insights into beta(2)-adrenergic receptor function. *Science* **2007,** *318*, 1266-1273.

(17) Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem Rev* **2017,** *117* (1), 139-155.

(18) El Hage, K.; Brickel, S.; Hermelin, S.; Gaulier, G.; Schmidt, C.; Bonacina, L.; van Keulen, S. C.; Bhattacharyya, S.; Chergui, M.; Hamm, P.; Rothlisberger, U.; Wolf, J. P.; Meuwly, M. Implications of short time scale dynamics on long time processes *Struct Dynam-Us* **2018,** *5*.

(19) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular signatures of G-protein-coupled receptors. *Nature* **2013,** *494*, 185-194.

(20) Massink, A.; Gutierrez-de-Teran, H.; Lenselink, E. B.; Zacarias, N. V. O.; Xia, L. Z.; Heitman, L. H.; Katritch, V.; Stevens, R. C.; IJzerman, A. P. Sodium Ion Binding Pocket Mutations and Adenosine A(2A) Receptor Function. *Mol Pharmacol* **2015,** *87*, 305-313.

(21) Dong, S. S.; Goddard, W. A.; Abrol, R. Conformational and Thermodynamic Landscape of GPCR Activation from Theory and Computation. *Biophys J* **2016,** *110*, 2618-2629.

(22) Backus, E. H. G.; Bloem, R.; Pfister, R.; Moretto, A.; Crisma, M.; Toniolo, C.; Hamm, P. Dynamical Transition in a Small Helical Peptide and Its Implication for Vibrational Energy Transport. *J Phys Chem B* **2009,** *113*, 13405-13409.

(23) Bogatyreva, N. S.; Ivankov, D. N. The relationship between the solvent-accessible surface area of a protein and the number of native contacts in its structure. *Mol Biol* **2008,** *42*, 932-938.