THE JOURNAL OF PHYSICAL CHEMISTRY B

COMMENT

pubs.acs.org/JPCB

Reply to "Comment on 'Toward Identification of the Reaction Coordinate Directly from the Transition State Ensemble Using the Kernel PCA Method'"

Dimitri Antoniou[†] and Steven D. Schwartz[*,†,‡,§]

[†]Department of Biophysics and [‡]Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, United States

[§]Institut des Hautes Études Scientifiques, 91440 Bures-sur-Yvette, France

The comment by Peters on our recent paper consists of misunderstandings regarding the algorithm used and of simple factual mistakes. We will address them below.

The first statement that Peters makes is that "competing methods" are more general. This is simply not true—the methods he references (either his own or related to his own) involve fitting of the committor. For his method to work, one needs a range of committor values. As we have stated in our manuscript, this is almost impossible when the process being studied is a direct chemical reaction with a rapidly rising committor. The only values of the committor reasonably obtained are zero or unity. In LDH, the rise occurs in less than 10 fs,[1] and it would be impossible to obtain a broad range of coordinate values to which one could fit the committor. The fitting methods, such as the work of Dinner and Ma[2] and the Trout group,[3] are far better suited to diffusive transition regions. Both his and our methods have their merits, but they are not "competing" since they are designed for different types of barrier crossing dynamics.

Peters then continues: "*Their strategy is not equivalent to choosing those residues that are involved in the mode of minimum TSE covariance. To identify the thinnest mode one should instead diagonalize the kPCA matrix.*" The above is a misunderstanding of the reason we employed the kPCA algorithm. We are interested in *directions* along which the separatrix is thin, not in the *thinnest* direction. A simple illustrative example: if a direction has length 10 and two other directions have length 1.1 and 1.0, we are indeed interested in both the latter directions and not only on the thinnest direction of length 1.0 that Peters' suggestion would identify. We know that in our system, similarly to the above simple example, one direction is much longer than the others because its corresponding eigenvalue dominates (it contributes 95% to the total variance).

Peters next says: "*Among the many bath modes—nearly three per atom—the proposed kPCA method identifies just one bath mode.*" This is wrong. The eigenvectors of kPCA are not related to an expansion around some energy minimum, so they are not bath modes. The kPCA eigenvectors are directions on the separatrix, not protein modes. Once we have the kPCA eigenvectors we go back to the trajectories of the TPS ensemble and examine the motions of residues that kPCA identified—it is only at this step that we identify dynamics. This confusion between the geometry of the separatrix and the dynamics of the protein leads to the next misunderstanding.

Peters continues: "*The use of a polynomial ($d = 2$) kernel to isolate one dominant bath mode in the kPCA method therefore lacks a theoretical justification. However, examining the TSE with standard linear PCA could reveal the mode that contributes the most to the activation entropy.*" These speculations are based on the erroneous identification of the geometry of the separatrix (as revealed by the kPCA eigenvector) with bath modes. In the original manuscript, we emphasized that we tried the linear PCA version and we explained why it did not work for this particular problem (the separatrix is far from a linear geometric object).

Next, the comment has a factual error: "*One might interpret from the authors' results on the lactate dehydrogenase (LDH) example that the chemical step is not important and therefore that the chemical step is not the dynamical bottleneck.*" This is simply not true: in every paper we have published on LDH we have clearly stated that the chemical step of hydride and proton transfer is included in the reaction coordinate. We have found that other degrees of freedom are necessary to define the reaction coordinate, as the reaction passes the transition state region. It is curious that the author of the comment is so sure that we imply that chemistry does not matter!

The next complaint Peters raises is: "*In contrast to the established procedure, Antoniou and Schwartz separately and simultaneously fix the positions of many different residues when sampling configurations for their histogram test.*" Once again, what the author of the comment calls "established procedure" refers to papers by Peters himself, but the established procedure is to fix a presumed reaction coordinate q to a value q^* that has commitment probability 1/2—if q is multidimensional, then the fixed value q^* is also multidimensional (e.g., see the review by Bolhuis, Chandler, Dellago, and Geisler,[4] pages 305−308). The author of the comment gives a toy counter-example, where if q is the true reaction coordinate, adding more constraints leads to an artificial overconstrained ensemble. This is correct. However, in that case the commitment probability $P(q^*)$ would have been 1/2, while in our enzymatic system when we constrained only few variables we were very far from $P = 1/2$; so, his toy counter-example is not relevant. To give our own toy example, consider a $P = 1/2$ separatrix that is defined by 2 degrees of freedom confined to 2 noncollinear lines: if one followed Peters' suggestion and held only one of those degrees of freedom fixed at a time and performed the constrained random walk to test the quality of a reaction coordinate, then one would clearly fall off the separatrix.

Peters next speculates: "*Therefore, an extra constraint on nearly any residue near the reaction center may narrow the committor*

Received: August 4, 2011
Revised: September 15, 2011
Published: September 19, 2011

distribution to some degree." Peters does not give his reasons for this speculation, but we have not seen it to be true in any of three enzymatic systems (LDH, PNP, DHFR (unpublished)) where we have applied TPS.

At the end, Peters makes a comment on our research program on rate-promoting motions: "*Computational methods should be developed and evaluated for their ability to impartially test hypotheses, and not for their ability to confirm a specific and controversial hypothesis.*[2]" What Peters calls a "*controversial hypothesis*", i.e., the effect of fast dynamics on catalysis, we call interesting. The methods we have used have rigorous mathematical underpinning as we have described in the original manuscript and in this response, and the problem to which we address these methods is, as Peters correctly suggests, at the forefront of modern computational biochemistry.

REFERENCES

(1) Quaytman, S.; Schwartz, S. D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12253–12258.
(2) Ma, A.; Dinner, A. R. *J. Phys. Chem. B* **2005**, *109*, 6769–6779.
(3) Peters, B.; Trout, B. L. *J. Chem. Phys.* **2006**, *125*, 054108. Peters, B.; Beckham, G. T.; Trout, B. L. *J. Chem. Phys.* **2007**, *127*, 034109.
(4) Bolhuis, P.; Chandler, D.; Dellago, C.; Geissler, P. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.