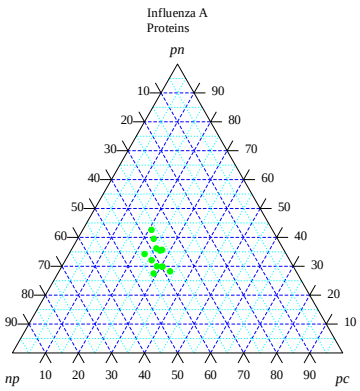
MOL2NET, International Conference Series on Multidisciplinary Sciences

Ternary Phase Diagrams of Viral Proteins: The Example of H1N1 Influenza

Yunjoo J. Bae and Daniel J. Graham*
Department of Chemistry and Biochemistry
Loyola University Chicago
Chicago, Illinois 60626 USA

*Corresponding Author: dgraha1@luc.edu

Graphical Abstract



Abstract.

The proteins encoded by a biological system present a curious mix of properties. On the one hand, the molecules are tasked with different functions: the system practices and depends critically on the division of labor. On the other hand, the proteins operate cooperatively: the tasks of one are complemented and, in multiple cases, brought to completion by another. How else could the system survive over the long term? The mix of diverse and cooperative traits can be explored with the help of thermodynamic tools. The variables of state for proteins can be coarse-grained so as to place points on phase diagrams. Variables which place points close to one another shine light on the cooperative facets of systems. Variables which disperse points shed light on the diversity of functions. There is one phase point for each protein so encoded and a proteome confers a locus of points. Here we illustrate a novel diagrammatic technique applied to viral systems, specifically influenza A. The motivation is to screen and identify the sites of proteins most critical to the cooperative and complementary nature. Such sites should be targeted for mutation or inhibition to attenuate the impact on host populations.

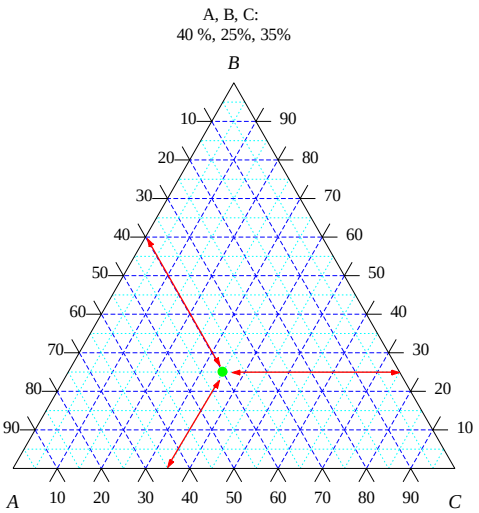
Introduction

The proteins encoded by a biological system present a curious mix of properties. On the one hand, the molecules are tasked with different functions: the system practices and depends critically on the division of labor. On the other hand, the proteins operate cooperatively: the tasks of one are complemented and, in multiple cases, brought to completion by another. How else could the system survive over the long term? Curiously, only the division of labor is apparent in the structures of proteins. Different functions require different amino acid sequences and 3D folds. The sequences and folds for system express sparse if any alignments.

The mix of diverse and cooperative traits can be explored with the help of thermodynamic tools. The variables of state for proteins can be coarse-grained so as to place points on phase diagrams [1, 2]. Variables which place points close to one another shine light on the cooperative facets of systems. Variables which disperse points shed light on the diversity of functions. There is one phase point for each protein so encoded and every proteome confers a locus of points. Here we illustrate a novel diagrammatic technique applied to viral systems, specifically influenza A. The motivation is to identify the sites of proteins most critical to the cooperative and complementary nature. Such sites should be targeted for mutation or inhibition to attenuate the impact on host populations.

Materials and Methods

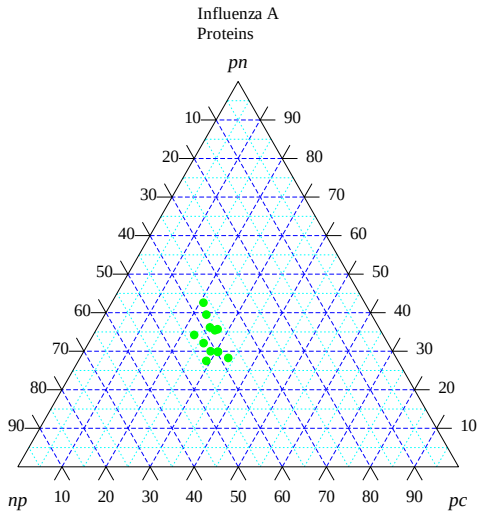
Ternary phase diagrams have roots in the nineteenth century [3]. They offer an eclectic way of projecting multi-dimension information onto a plane. It is easy to appreciate how they work, especially with components as state variables. Consider a three-component system *A*, *B*, *C* with respective mole fractions 0.400, 0.250, 0.350. The system places a state point in the equilateral triangle shown below. Note the red double-arrows tracing the point to percentages echoing the fractions. The graphing owes its success to a principle of equilateral triangles: lines perpendicular to each axis, and connecting with an internal point, are proportional to the percentages that place the point.



How do ternary plots accommodate proteins? There is a point placed for every protein encoded by a system. Where it places hinges on the component mole fractions *and* how the amino acids are partitioned. Different partitions offer different coarse-graining strategies and, in turn different perspectives. For example, we can partition the amino acids according to elementary side-chain functions:

Non-polar (*np*): A, V, L, I, P, F, W, M Polar Neutral (*pn*): G, S, T, C, Y, N, Q Polar Charged: (*pc*): D, E, K, R, H

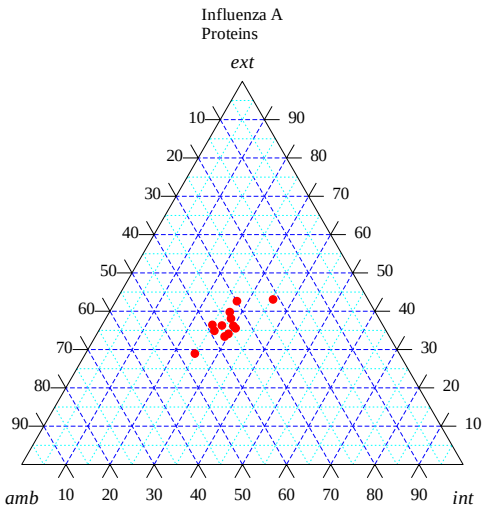
There are twelve proteins encoded by influenza A virus: polymerases, hemagglutinin, neuromidinase, etc. For an H1N1 Hong Kong strain of 1977, the phase plot appears below:



Another partition is guided by folding propensities:

External (*ex*): R, K, H, D, E, N, Q Ambivalent (*amb*): A, C, G, P, S, T, W, Y Internal (*int*): L, V, I, F, M

For the same strain, the plot appears below:



The two partitions offer different vantage points. And because they share identical measure distributions—the cardinality of the partition sets is 8, 7, 5—the distances between the phase points can be compared and contrasted. The first partition offers a more compact point locus compared with the second. The latter case was grounded on folding traits and reflects the diversity of protein tertiary structures. The point concentration

via the first attests to the sharing and overlap of side chain interactions of the components that render proteins.

Results and Discussion

Ternary phase plots provide insights when the partition space is explored. One seeks partitions that either minimize or maximize the distance between points. The phase diagrams of the previous section are based on solely on the component point of view. In contrast, minimum- and maximum-distance partitions are controlled by the system of interest. For example, the partitions

8_{min}: V, D, Q, K, F, A, S, Y

8_{max}: M, Q, K, W, T, H, R, L

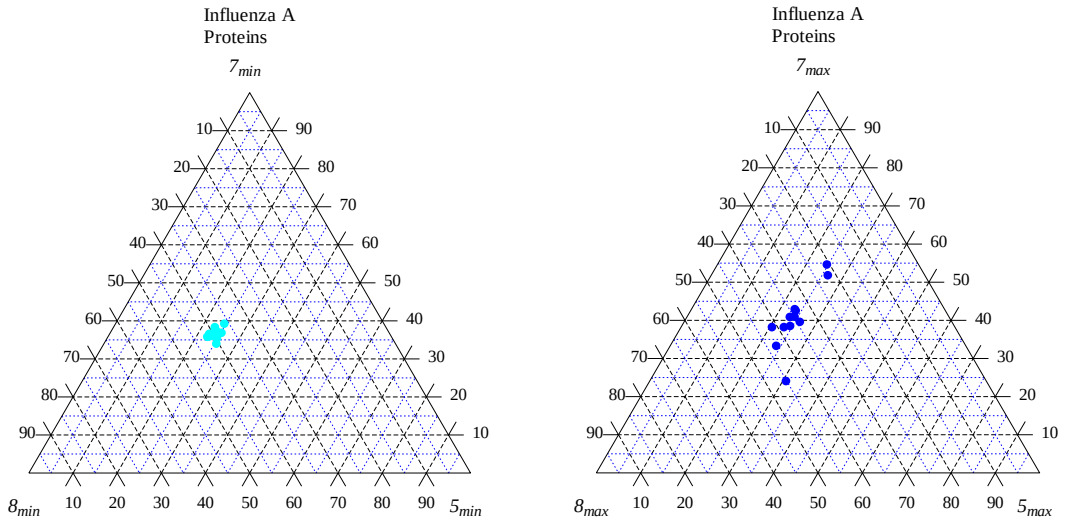
7_{min}: I, C, M, R, P, L, N

7_{max}: F, E, N, A, S, D, Y

5_{min}: E, W, H, G, T

5_{max}: P, V, I, G, C

underpin the following plots:



The left-side diagram shows a coalescence of the phase points whereas the right-side maximum dispersion and the trace of a curved phase boundary. By the left-side perspective, the proteins share intensive states whereby the gradients between states are minimized. In thermodynamic terms, the entropy production across proteins would be minimized by these intensive states. In the right-side view, the proteins express maximum divergence in intensive states: the gradients between states are accentuated; entropy production across proteins would be maximized by these intensive states. The potential for entropy production is a signature for cooperative and contrasting functions of proteins.

Conclusions

Multiple partitions achieve the contrasting and complementary effects of coalescence and dispersion of phase points. Insights arrive by identifying the components that do double-duty in significant ways: their placement in partitions has the effect of coalescing *or* dispersing the state points. For H1N1 influenza A,

these components are H, E, D, Y, and F. The cooperative nature and division-of-labor of proteins are critical to a system and present two sides of a coin. The identity of each site is dictated by evolution and its role in cooperative and diverse functions is obtuse. Explorations of partition space along with phase plots point to the sites most critical to the cooperative and division-of-labor balance. As stated previously, these present the most compelling targets for mutation and inhibition in order to attenuate the virulence. The case of influenza is an urgent one given epidemic threats and the absence of vaccinations.

References

[1] Graham, D. J., Grzetic, S, May, D., Zumpf, J. 2012. Information Properties of Naturally-Occurring Proteins: Fourier Analysis and Complexity Phase Plots, *Protein J.* 31, 550-563. doi:10.1007/s10930-012-9432-7.

[2] Graham, D. J. 2013. A New Bioinformatics Approach to Natural Protein Collections: Permutation Structure Contrasts of Viral and Cellular Systems, *Protein J.* doi:10.1007/s10930-013-9485-2.

[3] Laidler, K. J., Meiser, J. H., Sanctuary, B. C. 2003, Physical Chemistry, Houghton Mifflin, Boston, MA, Chapter Six.