# Protein Classification and Natural Kinds

Alexander W. Tolbert

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of
Master of Arts
In
Philosophy

Committee I. Benjamin Jantzen
Committee II. Lydia Patton
Committee III. Kelly Trogdon
Committee IV. Karen Kovaka

April 15, 2019
Blacksburg, Virginia

Keywords: Natural Kinds, Philosophy of Biology, Biochemical Kinds

# Protein Classification and Natural Kinds

Alexander W. Tolbert

## Abstract

*This project surveys biochemical practice and sets the record straight regarding which parts of protein classification are pluralist. Assuming an approach that attempts to draw metaphysical conclusions by analyzing how multifaceted practices of science work, I tie the results of my survey of protein classification practices to the debate over natural kinds. I address which classificatory practice is likely to pick out a natural kind. I defend the thesis that dynamics is a fundamental description of proteins as kinds. This view is widely held in biochemistry but is absent from philosophical literature on biochemical kinds.*

# Protein Classification and Natural Kinds

Alexander W. Tolbert

## Abstract

*This project surveys biochemical practice and sets the record straight regarding which parts of protein classification are pluralist. Assuming an approach that attempts to draw metaphysical conclusions by analyzing how multifaceted practices of science work, I tie the results of my survey of protein classification practices to the debate over natural kinds. I address which classificatory practice is likely to pick out a natural kind. I defend the thesis that dynamics is a fundamental description of proteins as kinds. This view is widely held in biochemistry but is absent from philosophical literature on biochemical kinds.*

Table of Contents

## I. Introduction

Biochemists recognize several general classes of macromolecules (nucleic acids, proteins, lipids, and carbohydrates) that are important for biological life. These scientific classes are referred to by philosophers as biochemical kinds (Bartol 2014, p. 531). There is a growing philosophical literature on biochemical kinds, especially proteins. Philosophers have begun to discuss the epistemic dimensions of how scientists classify proteins and to raise metaphysical questions concerning proteins' status as natural kinds.[1] One reason biochemical kinds are philosophically interesting is that they are a nexus between chemical elements and biological species. Like chemical elements, their physical structure is fundamental. But, as with biological objects, it is also important to understand biochemical kinds' physiological functions in addition to their physical structure (Bartol 2014, p. 531).

Bartol (2014) frames the philosophical literature on biochemical kinds as attempting to answer two distinct questions: the *epistemic question* and the *metaphysical question*. The epistemic question is, how can, do, and should scientists classify biochemical kinds? This question has normative and descriptive elements. The metaphysical question explores if proteins are natural kinds. That means is there a natural sorting of proteins into kinds, several such arrangements, or no such arrangement? The philosophical literature takes both monist and pluralist approaches to answer these questions. Monist approaches hold that there is one correct answer to the epistemic question and one correct answer to the metaphysical question. Pluralist approaches allow that these questions can have more than one correct answer.

[1] See (Slater 2009) (Tobin 2010) (Goodwin 2011) (Bartol 2014) (Havstad 2015) (Havstad 2018)

I address both the epistemic and metaphysical questions about biochemical kinds in this paper. My approach is informed by *scientific metaphysics*, the view that metaphysical inquiry into the structure of reality should be informed by the success of science (Ross et al. 2013) (Ladyman 2012). I argue that the epistemic question needs to be answered before the metaphysical question because a metaphysical view that is not credible with respect to the best contemporary scientific research should be doubted.

I begin by surveying biochemical practice and setting the record straight regarding which parts of protein classification are in fact pluralist. I build on Joyce Havstad's (2015) work because it provides the most detailed answer to the descriptive part of the epistemic question. Her assessment is that protein classification has three distinct classificatory practices (*characterization*, *individuation*, and *organization*). She claims that characterization and individuation are monist, while organization is pluralist. I build upon her account, amending it where necessary to reflect scientific practice. Most importantly, I show (contra Havstad) that when scientists classify proteins, they engage in only two distinct classificatory practices (*characterization* and *organization*), both of which are pluralist.

Next, I argue that the results of my survey of biochemical practice are relevant to the debate about natural kinds. Drawing on Ereshefsky and Reydon's (2015) natural kind theory, I show that the pluralism in biochemistry's classificatory practices is good evidence that the answer to the metaphysical question is also pluralist.

Finally, I conclude by gesturing at a positive proposal that argues that, by focusing primarily on biological and structural considerations, the scientific practice of protein classification has been grossly mischaracterized by philosophers' tendency to ignore dynamics, a fundamental paradigm in biochemistry (Henzler-Wildman and Kern 2007, p. 964). This

mischaracterization has consequences in that it has generated misguided metaphysical views that only see proteins as either chemical kinds, biological kinds or some combination. This project is novel because it considers protein dynamics as a fundamental description of proteins as kinds. This view is widely held in biochemistry but is largely absent in philosophical literature on biochemical kinds.

## II. Havstad's Tripartite Scheme

Havstad (2015) develops the most detailed answer to the descriptive aspect of the *epistemic question* by analyzing the scientific practice of protein classification. She claims that protein classification involves three distinct classificatory practices:

1. Characterizing the concept protein
2. Identifying which protein type a token protein belongs to
3. Sorting protein types into hierarchical classification schemes, or protein taxonomies

She argues that each of these classificatory practices is independent of one another. Thus, referring to protein classification in general as being monist or pluralist is not coherent. Instead one must specify a specific classificatory practice when making claims about monism or pluralism (Havstad 2015, p. 74). Crucially, each of these three classificatory practices invites questions about naturalness. That is, one can ask (a) whether the category "protein" is a natural kind, (b) whether protein types (e.g., RNA-polymerase II) are natural kinds, and (c) whether classes of protein types (e.g., kinases) are natural kinds.

I accept Havstad's claim that characterizing protein classification in general as being either monist or pluralist (instead of examining carefully a particular portion of the scientific practice in order to determine if this practice is monist or pluralist) makes little sense. But I reject Havstad's

tripartite picture of protein classification. Instead, I argue that there are two scientific practices of protein classification (*characterization* and *taxonomic organization*), both of which are pluralist.

## III. Critique of Havstad on Protein Characterization

In this section, I show that the concept *protein* has more than one meaning in scientific practice. The first type of classificatory practice in Havstad's tripartite scheme is what she calls *characterization*. Although she does not explicitly define protein characterization, it is clear she takes it to mean something definitional. She also believes that the classificatory practice of protein characterization is monist: "Although there is significant scientific concern surrounding the characterization of the concept *species*, for instance, there is no such controversy surrounding the characterization of the concept *protein*" (Havstad 2015, p. 74-75).

Havstad is correct that there is not much scientific literature debating the concept *protein*, but this fact does not mean there is a clear consensus regarding the concept in the scientific community. On the contrary, Mannige (2014) describes two different characterizations of the concept *protein* that exists in the scientific literature and states there is no clear consensus among biochemists about which characterization should be used.

Mannige begins with the canonical definition or characterization of protein present in many textbooks. This characterization—henceforth called the ***narrow view***—recognizes a protein as a (>50 amino acid) series of peptides that possess the ability to adopt one or several conformations. Because of the long-standing belief that a protein first adopts a particular conformation and then takes on a biological activity, the view persisted for years that proteins are differentiated from

peptides (i.e., peptide chains that are not proteins) chiefly by their ability to fold and by their longer lengths (Mannige 2014, p. 133).

Mannige then describes a second characterization, which he calls the ***expanded view.*** Mannige points out that small folded protein domains (such as the zinc finger domain) and dynamic proteins are recognized by some biochemists as constituting a considerable part of the proteome (the entire set of proteins that can be expressed by a genome). These cases have compelled some scientists to reconsider the difference between protein and peptide. Mannige argues that the narrow distinction does not capture the true nature of the entire proteome. He suggests expanding the older conception to recognize a protein as being a peptide chain that can adopt one or several discrete conformations or binds consistently to at least one particular cognate partner. The expanded view classifies dynamic, frequently disordered, yet still functional, peptide chains—common entities in the proteome—as genuine proteins (Mannige 2014, p. 133).

According to Mannige, there is no consensus in biochemistry about whether the narrow or the expanded characterization of proteins is correct (Mannige 2014, p. 133). This example shows that scientific classificatory practice uses more than one characterization of what proteins are. Therefore, Havstad is wrong to describe this practice as monist. Instead, it is pluralist: with respect to the question *what are proteins?* scientific practice does permit several characterizations.

## IV. Critique of Havstad on Protein Individuation and Organization

The final two pieces of Havstad's schema are protein *individuation* and *organization*. According to Havstad, *individuation* is the sorting of individual token proteins into maximally specific protein types or kinds according to their genetic origin, protein superstructure, amino acid

sequence, and species specificity. She asserts that there is a scientific consensus regarding protein individuation. That is, there is one system for typing proteins according to which each token protein molecule is a member of one and only one protein type, from the point of folding throughout its dynamic life until denaturation (Havstad 2015, p. 75). For example, scientists seem to agree about which molecules belong in the extension of the term "titin." This agreement shows that the practice of protein classification is monist.

*Organization*, on the other hand, consists of the variety of ways of arranging token proteins into taxonomies, according to their properties. Provided that individuation is monist, this really means arranging the protein types used for individuation into meta-classes, and meta-meta-classes, etc. According to Havstad, token proteins are complex in that they possess several properties and capacities that interact and cluster in different ways. As a result of this kind of complexity, there are several ways to sort or organize protein types, and scientists do in fact make different choices in different contexts. This means that protein organization is pluralist: "Just to be clear, I'm saying that, although everyone agrees on the individuation of the basic entities being sorted, there is a variety of ways of arranging these entities into hierarchical classification schemes, or taxonomies. And this is all within the relevant scientific field" (Havstad 2015, p. 76).

But is Havstad right about individuation? That is, does there exist a unique, fine-grained classification of molecules or parts of molecules into protein types such that in all contexts, scientists would agree that molecular or sub-molecular structure $x$ is a token of protein type $T$? Havstad is correct that answering this claim requires scrutiny of protein classification practice.

It is not clear to what specific scientific practice Havstad is referring when she claims that classificatory practice is unified regarding protein individuation into "basic entities" belonging to minimal protein types. Havstad suggests that one salient membership criterion for protein individuation is superstructure. But this requires establishing definitions of distinct, smaller structural units that comprise the superstructure itself. In biochemical practice, there are sets of definitions for these smaller units, spanning a range of scales of length and structure (Hadley and Jones 1999, p. 1099–1100). Each set of definitions would yield different individuation criteria, incompatible with the others. In other words, different structural unit definitions corresponding to different classification systems yield different judgments about superstructure (Hadley and Jones 1999, p. 1111). The work of Hadley and Jones (1999) shows how the definition of structural units is somewhat arbitrary and determined by the interests of the creators of classification methods. They show how the subjective nature of structural definitions leads to different definitions of structural units across different classification databases. This results in differences in how proteins are sorted across distinct classification systems. Similar research also demonstrates that different rules, particularly with respect to structural definitions, can result in different arrangements of the same protein or molecular subcomponents (Csaba et al. 2009, p. 10, 18).

Consider the following example. SCOP (Structural Classification of Proteins) and CATH (Class Architecture Topology Homology) are the two most pre-eminent protein structure classification systems and are considered the gold standard (Csaba et al. 2009, p. 10, 16). They are in some sense sub-molecular classification schemes in that, in order to name a molecule, both schemes first partition it into smaller named components whose identity and arrangement determine the classification of the overall molecule. Proteins are described at several levels of structure: primary structure (linear sequence of amino acids), secondary structure, (stable recurring

geometrical patterns), tertiary structure (geometric and bond structures of the whole molecule), and quaternary structure (the molecule plus bound partners) (Bartol 2014, p. 535). Protein structures are regularly comprised of distinct globular *domains* -- compact, local, partially independent folding sections formed from secondary structures like helices and sheets. Demarcation and identification of domains within protein structures are challenging and dependent on the choice of rules deployed. There are several algorithms that exist that are capable of domain assignment however, each uses a different set of rules. *Class* is commonly defined by the total arrangement of secondary structural components within a domain. A *fold* is defined by the number, arrangement, connectivity, and topology of secondary structure. A *superfamily* is a set of domains with comparable folds (and often related functions which suggest common ancestry despite lacking similarity of amino acid sequence) (Hadley and Jones 1999, p. 1099–1100).

SCOP and CATH use different rules which sometimes categorize the same protein sample differently. For example, consider the protein papain. According to SCOP, this protein has a single domain. SCOP leaves the catalytic amino acids together to form the active site. CATH, on the other hand, divides papain into two domains, dividing the cysteine from the asparagine and histidine. This assignment leaves each domain without function. Usually, CATH assigns more domains than SCOP. This is because CATH uses a structure based definition for domains. SCOP, on the other hand, considers if a domain is recognized as reappearing in another superfamily, or recognized as a distinct single-domain fold (Hadley and Jones 1999, p. 1104–1105). The other units (class and fold) of the multilayer systems have similar complications.[2]

[2] See (Hadley and Jones 1999, p. 1104–1107)

Another difficulty is that while it is commonly thought that a protein in its native state assumes a single, well-defined conformation, there are some instances that show the same amino acid sequence can fold multiple ways (Fox et al. 1986, p. 192, 194). Intrinsically disordered proteins (IDPs) are examples of proteins that can have the same amino acid sequence but can assume different structures. Some prion proteins adopt different structures with the same sequence because of misfolding. The same sequence can fold uniquely and this furnishes distinct structures with potentially different biological functions (Moreno-Gonzalez and Soto 2011, p. 1–3). IDPs are proteins that lack a definite three-dimensional structure. IDPs have many conformational states and active conformational dynamics (Mahmoudabadi et al. 2013, p. 26–27).

Not only can the same amino acid sequence take on different structures, but *different amino acid sequences* can also be classified as the *same* protein because the superstructure is essentially the same (Friedberg and Margalit 2002, p. 350). This is because the superstructure of a protein is typically robust against mutations (Sikosek and Chan 2014, p. 17). This lack of a one-to-one mapping amino acid sequence and structure means that Havstad is wrong: appealing to superstructure or primary structure is not sufficient for protein individuation.

## V. Aim and Individuation: Dynamic Classifications

This section argues two further points. First, Havstad's separation between organization and individuation is a distinction without a difference when we consider scientific practice. Second, protein dynamics (that is, protein *motion*) are an important dimension of protein classification.

The previous section showed that Havstad is mistaken when she asserts there is a consensus among biochemists about how to generate lists of entities that belong in particular groups. However, Havstad's work does get at the second fundamental question protein classification practice is interested in answering: whether or not two protein samples are the same kind. This

question is ambiguous if asked in isolation, but it is made precise largely by the aims and interest of researchers. Hensen et al. (2012) state: "The core problem of any classification approach is the choice of a proper metric, which discerns similar from different samples" (p. 2). That is, scientists determine whether two proteins are of the same kind by first choosing a given metric then by examining the properties of tokens. There are several organization schemes for generating lists of entities that belong in particular groups, but none seem to suggest any sort of consensus or fundamentality (Hadley and Jones 1999, p. 1111). Different taxonomic classifications represent different epistemic aims of biochemistry. There are at least three such classifications: structural classifications, functional classifications, and dynamic classifications.[3]

The idea that dynamics is one of the fundamental characteristics of proteins has been largely ignored in the biochemical kinds literature. However, researchers have begun to classify proteins in dynamical space. Hensen et al. (2012), for example, show that proteins fall into dynamical classes. Proteins are always in motion. This motion is called dynamics. Dynamics is jointly determined by the aggregate impact of the forces on and exerted by intrinsic factors (all atoms that constitute the protein) and extrinsic factors (environmental conditions) (van der Kamp et al. 2010, p. 423). A comprehensive characterization of proteins necessitates a dynamic multidimensional energy landscape which determines the relative probabilities of the conformational states and the energy barriers among them. This view led to an expansion of the sequence-structure-function model to incorporate dynamics. Understanding proteins' biological activity requires time to be added to the still-frames of proteins in crystal structures (Henzler-Wildman and Kern 2007, p. 964). While three-dimensional protein structures provide some understanding, alone they cannot

[3] Protein can also be sorted into groups based on function which are cataloged by the Gene Ontology (GO) (Natale et al. 2011).

account for the observed properties and functions in any mechanistic sense (Debrunner and Frauenfelder 1982, p. 283). Thus, understanding the structure-function relationship of proteins requires examining their dynamics as well as their structure.

Several studies have shown that the connection between structure and function does not, when lacking information about sequence similarity, permit for strong function predictions. Structures that are highly different can have the same function and highly comparable structures can have vastly distinct functions. For example, the TIM-barrel fold has been identified with almost all enzymatic activity. Consequently, structure-based protein function predictions have not been successful at predicting protein function exceeding 30% fidelity (Hensen et al. 2012, p. 2). Thus, dynamic-based databases are a valuable tool for biochemists and any account of proteins that ignores dynamics has mischaracterized the fundamental role dynamics plays in describing proteins in scientific practice.

The upshot of these considerations is twofold. First, even if we assume that superstructure is one of the primary criteria for individuation (as Havstad suggests), this is ambiguous because there are several options for defining structural units based on which schema we adopt. Second, in practice, the distinction between individuation and organization falls apart. That is, there is no unique structural level at which parts we might call proteins are individuated (submolecular, molecular, multi-molecule units, cross-cutting units like active sites), meaning that questions of protein individuation just are questions of protein organization, leading us back to my first point. There are several classification schemes depending on aims. Depending on what you are classifying for, you will adopt one or another principle of individuation. Because there are different ways of classifying the same material parts, it is hard to see that individuation is any different from

organization. What we really have is a general problem of individuating biological molecules, and there is no one answer in practice. No matter where you start, you end up with pluralism.

## VI. Scientific Metaphysics as a Guide to Inform Biochemical Kinds.

In this section, I tie the results of my survey of protein classification practices to the debate over natural kinds. In particular, I address which classificatory practice is likely to pick out a natural kind.

### *Natural vs Scientific Kinds*

Historically there are two philosophical debates that go under the title of "the natural kind debate." The first debate is about whether there are any kinds out there that exist independently of human classificatory activities. That will not be my focus here. The second natural kind debate originates with Nelson Goodman's new riddle of induction. Goodman formulated a powerful new version of skepticism about our ability to make justified inductions. This skepticism arises from underdetermination in the classification schemes we do or can use to make inductions (e.g. green, red, & emerald v. grue, gred, & emeruby) (Goodman 1983). Quine's proposed solution to Goodman's riddle (in *Ontological Relativity and Other Essays*) was that the only justified inductions are those made using natural kinds (Quine 1969). Eventually, Goodman adopted Quine's response to his problem (see *Ways of Worldmaking),* but what a natural kind is beyond those kinds that allow us to make reliable inductions has been elusive to philosophers (Goodman 1978).

In the tradition of Goodman kind-theorizing, Ereshefsky and Reydon, (2015) offer a promising view of kinds that attempts to answer the original epistemic concern that Goodman's riddle raised. In their account, a domain of science taxonomic system constitutes a natural kind when (a) the

sorting principles of a taxonomic system support that classification's motivating principles, (b) the motivating or sorting principles of a taxonomic system are empirically testable, and (c) the taxonomic system is not degenerative with respect to other rival taxonomies.

Ereshefsky and Reydon's three criteria are not intended to be authoritative for discovering natural kinds. Their major presumption is that the best theories and classificatory programs the sciences have to offer is the ideal location to discover natural kinds. Still, it is possible that current scientific theories and programs can be mistaken. According to their view, natural kinds are primarily categories identified by the best theories and classification systems of science. They are not necessarily eternal ontological categories (Ereshefsky and Reydon 2015, p. 984). Here, I apply Ereshefsky and Reydon's criteria to protein classification practices and ask if any of the resulting taxonomies pick out natural kinds.

*Internal Consistency of Kinds*

Ereshefsky and Reydon's first criterion necessitates internal coherence. For any scientific discipline, methodological rules should be chosen that support the goals of that domain (Ereshefsky and Reydon 2015, p. 980). For them, the sorting principles of a taxonomic system should advance that system's motivating principles. Sorting principles organize objects into classifications within a taxonomy. Motivating principles provide warrant for the application of sorting principles (Ereshefsky and Reydon 2015, p. 979). Failure to meet this criterion means that the kinds obtained by a set of sorting principles are without justification with regard to the aims of the taxonomic system. A natural kind should be non-arbitrary. Therefore, internal consistency is a crucial requirement for a classificatory system to pick out natural kinds (Ereshefsky and Reydon 2015, p. 980).

Ereshefsky and Reydon assert that the Biological Species Concept is an illustration of a taxonomic system that meets the internal coherence requirement. They examine the Biological Species Concept's sorting principles and motivating principles. The Biological Species Concept's motivating principle is the thesis that interbreeding and the existence of isolated gene pools produce stable and discrete collections of organisms. The Biological Species Concept's sorting principle views organisms that have the ability to interbreed as identical species, organisms that cannot interbreed are considered different species, and organisms that breed asexually are not considered species. The Biological Species Concept's sorting principles fulfill the taxonomic system's goals. As a result of these principles populations that interbreed are reproductively separated from different collections and are a stable taxonomic group in evolution. The Biological Species Concept satisfies the conditions that a taxonomic system's sorting principles promote its motivating principles (Ereshefsky and Reydon 2015, p. 979-980).

Turning back to biochemistry, the sorting principles of the practices of protein characterization and organization also satisfy this requirement. The goal of these practices is not to provide a uniquely correct account of the classification of macromolecules. Instead, the motivating principle is to create consistent and identifiable classificatory units. The sorting principles of protein characterization use structural and dynamic definitions to characterize protein kinds (Mannige 2014, p. 133). The sorting principles of protein organization use several definitions and comparative metrics based on structure, dynamics, and function to produce different protein

taxonomies.[4] Both of these sets of principles satisfy their programs' aims by producing recognizable and stable taxa.

We can contrast this with a case that Ereshefsky and Reydon provide as an example that fails to meet their criteria. They consider the Phenetic classification system in biology. This system's goal is to generate classifications that offer no theoretical presuppositions. The chief sorting principle of the Phenetic classification is to organize organisms with respect to their total similarity. The difficulty for Phenetics is that its motivating principle is not capable of being achieved by way of its sorting principle. This is because organisms have an indeterminate amount of related features, consequently, some similarities have to be chosen for generating a taxonomy. Pheneticists must apply theoretical choices to select which characteristics to utilize for generating classifications. Thus, we see that pheneticist's motivating principle is to avoid theoretical assumptions and its sorting principles requires invoking theoretical considerations in order to select which traits generate classifications (Ereshefsky and Reydon 2015, p. 980-981). Due to these facts, Phenetics' sorting principles stop the achievement of its motivating principle. Therefore, Phenetics is not a consistent taxonomic system does not constitute a natural kind.

*Testability of Kinds*

Ereshefsky and Reydon 's second requirement is that the motivating or sorting principles of a taxonomic system should be testable empirically. If these principles are not empirically testable then there is no means of judging if they have any correlation with the empirical universe. This requirement for empirical testability is consistent with the belief that natural kinds are to some

[4] See (Hadley and Jones, 1999) (Pal 2006) (Hensen, et al., 2012) (van der Kamp et al. 2010)

degree based in nature. That is, natural kinds are collections in the empirical universe that are, to some extent, independent of human classificatory interest and aims. For example, Ereshefsky and Reydon believe the Biological Species Concept meets this criterion due to the fact the program's motivating principles present assertions regarding the empirical world and these claims have been tested extensively empirically (Ereshefsky and Reydon 2015, p. 981).

The sorting principles of protein characterization and organization are empirically testable, just as the sorting principles of the Biological Species Concept are. The goal of both characterization and organization is to find stable and recognizable categories. When biochemists characterize the concept of protein and organize proteins into taxonomies according to particular criteria (e.g. microstructure, dynamics, superstructure, biological activity, etc.) scientists are capable of examining if the data they have collected is about those organizational metrics. For example, inositol hexakisphosphate kinases (IP6Ks) are sorted into a group of proteins called kinases (an enzyme that modifies other molecules by chemically adding phosphate groups to them). These molecules have three mammalian isoforms. IP6K2 (one of these isoforms) is able to be tested by biochemists through various chemical assays to see if it has kinase activity (Wang et al. 2014, p. 10). The information collected is relevant to the organization metric scientists are paying attention to, which, in this case, is kinase activity.

*Progressiveness of Kinds*

Ereshefsky and Reydon assert that although internal consistency and empirical testability are required standards for classificatory programs to constitute natural kinds, they are not both necessary and sufficient. For example, phlogiston theory is a collapsed taxonomic system but still

an empirically testable classification system. Due to examples like this, an additional restriction is needed for deciding which taxonomic systems pick out natural kinds (Ereshefsky and Reydon 2015, p. 982). For them, a classificatory program must also be progressive. A taxonomic system is progressive if it furnishes rules capable of generating further classifications or increases existing classifications that are empirically successful with respect to opposing taxonomic programs (Ereshefsky and Reydon 2015, p. 982).

The classificatory practices of protein characterization and protein taxonomy are examples of progressive classificatory programs. Their sorting rules allow biochemists to create consistent and identifiable classifications of proteins. They have led to many new flourishing taxonomies, as well as contributed to the abolition of old concepts of protein and the refinement of current concepts (Stretton 2002, p. 527–528) (Mannige 2014, p. 133). Structure-based protein classification is crucial for gaining knowledge about the connection between protein structure and protein sequence evolution. SCOP and CATH, which are gold standards in the biochemistry for protein classification, are used in several machine learning methods and assessments of structure prediction (Csaba et al., 2009, p. 9-10, 16). Often differences are compared across databases as a means to expand and refine knowledge of proteins (Csaba et al., 2009, p. 17). This often leads to the generation of new databases with different aims and metrics such as the dynamic classifications mentioned earlier (Andreeva et al. 2014, p. 310–311, 313) (Hensen, et al., 2012, p. 1-2). Given the prior assessment, if we take Ereshefsky and Reydon's natural kind theory to be a plausible of account of kinds, then the protein classificatory practices that are likely to pick out natural kinds include both characterization and organization. This means that dynamic classifications likely yield natural kinds—a view that has for the most part been absent from biochemical kind literature.

## VII. Conclusion

The strategy of focusing on scientific practice in order to improve the results of metaphysics has recently been recently pursued by Love and Brigandt (2017) with respect to the metaphysics of biological individuality. Here I have used the same strategy to address metaphysical questions concerning the status of proteins as natural kinds. I have attempted to set the record straight regarding what types of protein classifications exist in biochemical practice and whether they are monist or pluralist. I tied the results of this survey to the debate over natural kinds. Relying on Ereshefsky and Reydon's natural kind theory because of its empirical adequacy considerations, I determined that both protein characterization and organization satisfy their natural kind theory.

Department of Philosophy

Virginia Tech University

Blacksburg, VA

altol25@vt.edu

# References

Andreeva, A. et al., 2014. SCOP2 prototype: a new approach to protein structure mining. *Nucleic acids research*, 42(Database issue), pp.D310–4.

Bartol, J., 2014. Biochemical Kinds. *The British journal for the philosophy of science*, 67(2), pp.531–551.

Csaba, G., Birzele, F. & Zimmer, R., 2009. Systematic comparison of SCOP and CATH: a new

gold standard for protein structure analysis. *BMC structural biology*, 9, p.23.

Debrunner, P.G. & Frauenfelder, H., 1982. Dynamics of Proteins. *Annual review of physical chemistry*, 33(1), pp.283–299.

Ereshefsky, M. & Reydon, T.A.C., 2015. Scientific kinds. *Philosophical studies*, 172(4), pp.969–986.

Fox, R.O., Evans, P.A. & Dobson, C.M., 1986. Multiple conformations of a protein demonstrated by magnetization transfer NMR spectroscopy. *Nature*, 320(6058), pp.192–194.

Friedberg, I. & Margalit, H., 2002. Persistently conserved positions in structurally similar, sequence dissimilar proteins: roles in preserving protein fold and function. *Protein science: a publication of the Protein Society*, 11(2), pp.350–360.

Goodman, N., 1983. *Fact, Fiction, and Forecast*, Harvard University Press.

Goodman, N., 1978. *Ways of Worldmaking*, Hackett.

Goodwin, W., 2011. Structure, function, and protein taxonomy. *Biology & philosophy*, 26(4), pp.533–545.

Hadley, C. & Jones, D.T., 1999. A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. *Structure* , 7(9), pp.1099–1112.

Havstad, J., 2015. Protein tokens, types, and taxa. In C. Kendig, ed. *Natural Kinds and Classification in Scientific Practice*. Routledge, pp. 94–106.

Havstad, J.C., 2018. Messy Chemical Kinds. *The British journal for the philosophy of science*, 69(3), pp.719–743.

Hensen, U. et al., 2012. Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. *PloS one*, 7(5), p.e33931.

Henzler-Wildman, K. & Kern, D., 2007. Dynamic personalities of proteins. *Nature*, 450(7172), pp.964–972.

van der Kamp, M.W. et al., 2010. Dynameomics: a comprehensive database of protein dynamics. *Structure* , 18(4), pp.423–435.

Ladyman, J., 2012. Science, metaphysics and method. *Philosophical studies*, 160(1), pp.31–51.

Love, A.C. & Brigandt, I., 2017. Philosophical dimensions of individuality. In S. Lidgard & L. K. Nyhart, eds. *Biological Individuality: Integrating Scientific, Philosophical, and Historical Perspectives*. University of Chicago Press, pp. 318–348.

Mahmoudabadi, G. et al., 2013. Intrinsically disordered proteins and conformational noise:

implications in cancer. *Cell cycle* , 12(1), pp.26–31.

Mannige, R.V., 2014. Dynamic New World: Refining Our View of Protein Structure, Function and Evolution. *Proteomes*, 2(1), pp.128–153.

Moreno-Gonzalez, I. & Soto, C., 2011. Misfolded protein aggregates: mechanisms, structures and potential for disease transmission. *Seminars in cell & developmental biology*, 22(5), pp.482–487.

Natale, D.A. et al., 2011. The Protein Ontology: a structured representation of protein forms and complexes. *Nucleic acids research*, 39(Database issue), pp.D539–45.

Pal, D., 2006. On gene ontology and function annotation. *Bioinformation*, 1(1), pp.97–98.

Quine, W.V.O., 1969. *Ontological Relativity and Other Essays*, Columbia University Press.

Ross, D., Ladyman, J. & Kincaid, H., 2013. *Scientific Metaphysics*, Oxford University Press.

Sikosek, T. & Chan, H.S., 2014. Biophysics of protein evolution and evolutionary protein biophysics. *Journal of the Royal Society, Interface / the Royal Society*, 11(100), pp.20140419–20140419.

Slater, M.H., 2009. Macromolecular Pluralism. *Philosophy of science*, 76(5), pp.851–863.

Stretton, A.O.W., 2002. The first sequence. Fred Sanger and insulin. *Genetics*, 162(2), pp.527–532.

Tobin, E., 2010. Microstructuralism and macromolecules: the case of moonlighting proteins. *Foundations of Chemistry*, 12(1), pp.41–54.

Wang, H. et al., 2014. IP6K structure and the molecular determinants of catalytic specificity in an inositol phosphate kinase family. *Nature communications*, 5(1). Available at: http://dx.doi.org/10.1038/ncomms5178.