Complexity, BioComplexity, the Connectionist Conjecture and Ontology of Complexity

Debaprasad Mukherjee

School of Information Technology,
Bengal Engineering and Science University, Shibpur
Howrah, West Bengal, India 711103

email: biodeb@gmail.com

Keywords: Complexity; Totality; Conjecture; BioComplexity; Connectionism; Network; Ontology; Information; Computation; Randomness; Evolution; Emergence; Criticality; Symmetry; Taxon; Species; Signaling; Communication; Control; Protein; Families; Domains; Algorithm; Clustering; Dimensions; Spaces; Nonlinearity; Methodology

Abstract:

This paper develops and integrates major ideas and concepts on complexity and biocomplexity - the connectionist conjecture, universal ontology of complexity, irreducible complexity of totality & inherent randomness, perpetual evolution of information, emergence of criticality and equivalence of symmetry & complexity. This paper introduces the Connectionist Conjecture which states that the one and only representation of Totality is the connectionist one i.e. in terms of nodes and edges. This paper also introduces an idea of Universal Ontology of Complexity and develops concepts in that direction. The paper also develops ideas and concepts on the perpetual evolution of information, irreducibility and computability of totality, all in the context of the Connectionist Conjecture. The paper indicates that the control and communication are the prime functionals that are responsible for the symmetry and complexity of complex phenomenon. The paper takes the stand that the phenomenon of life (including its evolution) is probably the nearest to what we can describe with the term “complexity”. The paper also assumes that signaling and communication within the living world and of the living world with the environment creates the connectionist structure of the biocomplexity. With life and its evolution as the substrate, the paper develops ideas towards the ontology of complexity. The paper introduces new complexity theoretic interpretations of fundamental biomolecular parameters. The paper also develops ideas on the methodology to determine the complexity of “true” complex phenomena.

Introduction

Complexity is a field of study whose fundamental purpose is to understand the incomprehensible or the undefined. There are several definitions of the field but non are significantly accepted. In on aspect, the field deals with finding the common laws across different systems, processes or phenomena at various scales of abstraction and projections, both in the natural and the artificial world. In one pragmatic sense, complexity includes the study of complex systems, esp. those which are composed of a number and variety of parts. Biological systems can be categorized as complex systems and the field which studies the complexity of biological systems is termed as “BioComplexity”. The process of evolution of life is considered as one factor that contributes to the complexity of life in its entirety.

Complexity of Bio-systems and Dynamical Evolution: Because of evolution and a long history of environmental accidents that have driven processes of natural selection, biological systems can be regarded as engineered entities than as objects whose existence might be predicted on the basis of the first principles of physics, although the evolutionary context means that an artifact is never “finished” and rather has to be evaluated on a continuous basis. The task of understanding biological systems thus becomes, to a significant extent, one of “reverse engineering”—attempting to understand the construction of a device about whose design little is known but from which much indicative empirical data can be extracted. In the highly interactive systems of living organisms, the macromolecular, cellular, and physiological processes, themselves at different levels of organizational complexity, have both temporal and spatial components. Interactions occur between sets of similar objects, such as two genes, and between dissimilar objects, such as genes and their environment. A key aspect of biological complexity is the role of chance. One of the most salient instances of chance in biology is evolution, in which chance events affect the fidelity of genetic transmission from one generation to the next. The hand of chance is also seen in the development of an organism—chance events affect many of the details of development. But perhaps one of the most striking manifestation is that individual biological organisms—even as closely related as sibling cells—are unlikely to be identical because of stochastic events from environmental input to thermal noise that affect molecular-level processes. If so, no two cells will have identical macromolecular content, and the dynamic structure and function of the macromolecules in one cell will never be the same as even a sibling cell. Example of biological complexity is illustrated by the fact that levels of biological complexity extend beyond the intricacies of the genome and protein structures through supramolecular complexes and organelles to cellular subsystems and assemblies of these to form often functionally polarized cells that together contribute to tissue form and function and, thereby to an organism’s properties. Complexity, then, appears to be an essential aspect of biological phenomena. Thus modern biology needs to integrate a number of diverse intellectual themes e.g. that of the reductionist and systems approaches. The processes of biology, the activities of living organisms, involve the usage, maintenance, dissemination, transformation or transduction, replication, and transmittal of information across generations. Biological systems are characterized by individuality, contingency, historicity, and high digital information content—which can be summarized by saying “every living thing is unique”. Furthermore, the uniqueness and historical contingency of life means that for population-scale problems, the potential state space that the population actually inhabits is huge. The study of such unique entities or phenomenon demands the integration of computing into biological research. But this integration of computing into biological research raises epistemological questions about the nature of biology itself. The question of how biological phenomena can be understood has not been fully settled. The behavior of a given system is explained to an extent by the behaviors of the components that make up that system. In addition to understanding its individual components, understanding of a biological system must also include an understanding of the specific architecture and arrangement of the system’s components and the interactions among them.

Every system or phenomenon undergoes changes in its design or dynamics, either naturally or artificially with time or in generalized phase space appropriate for the context. This process of change can be termed as evolution. Biological systems or phenomenon as it is seen today has also been subjected to such processes of evolution and probably has achieved such unique and fascinating characteristics due to the nature of the evolutionary process. According to one of the most accepted view of evolution, the process of evolution is a kind of multi-objective optimization process where inherent random noise percolates through the biosphere and influences the trajectory, to a significant extent, in the phase space of solutions of stable yet open systems i.e. individual organisms. The role the entire biosphere plays in this complex evolutionary process is observable through the constraints set by the bounded yet changing environment and the network dynamics of the organisms within and across themselves to create the ecosystem, which, in turn, affects the process of evolution- a cyclic and recursive phenomenon. The balance and interplay between the processes of mutation, viewed at different levels of abstraction and the selection and random drift creates a scenario where unity within diversity or "equivalently" diversity within unity, either way, is the natural state of the biosphere. Selection can be either positive or negative, where entities with beneficial properties are propagated through the ecosystem or entities with deleterious properties slowly become extinct. Although evolution is partially synonym with the history of life on earth, yet it must be taken into consideration that the present is not just a continuation of the past, but rather emergent from the past. The mix and match process of biological evolution traversing a complex multidimensional phase space in search of optimal or minimally optimal solutions inherently destroys the conventional statistical nature of the biosphere at various levels of abstractions and triggers the emergence of network behaviors. These networks can respond to external or internal perturbations i.e. either information or noise, as a whole and maintain the dynamic stability (robustness) of the entities e.g. organisms. Yet these networks allow the organisms to execute the inherent property of adaptability within the complex uncertainty of the environmental constraints or demands.

The processes of genome evolution which acts on these genomes are wide, varied and fascinating. The fundamental processes of molecular evolution influence the evolution of the genome and vice versa i.e. the evolution of the genome as a whole often influences the evolution of single genes or gene families since the entire genome acts a whole network, where one component influences the functioning of another. Genome phylogeny is an important method to delineate the evolutionary history of a set of genomes. But there are several problems delineating such phylogeny due to the complex dynamics of genomes especially through exchange of genetic material or genome segments through external sources. This is also due to the internal dynamics of the genome, which constantly reorders the chromosomal synteny on an evolutionary time scale.

The initial seed bias at any point of evolutionary time in genomes for its various state variables induces the evolutionary dynamics within the population spectrum of various species and subspecies. This, in turn, changes the genome design and consequently the intra-genome dynamics of the individual organism (e.g. exchange of functional genetic material and thus the re-adjustment of the biomolecular networks or pathways to return to the stable regime which was perturbed due to integration of foreign genetic material). Such changes in intra-genome dynamics, in turn, influences the evolutionary processes at the population level, which again influence the intra-genome dynamics at the individual organisms level. This cycle goes on shaping the fascinating evolutionary process of the complex genomes. The speciation process resulting in the diversification of new species from ancestral species generates innovations in the genomes influencing the morphology and behavior of the organisms/species. These vertical (e.g. speciation) and horizontal genetic innovations and sometimes discontinous changes in the environments have shaped the genomes in the living kingdom including functional aspects of organisms such as signaling and regulation.

Ab-intio (with no a priori biological hypothesis) global studies of genomes has the potential to elucidate the organizational structure of such genomic or functional systems (e.g. at the molecular level). These studies leads to the actual characterization and the nature of the variations within and across genomes, thus enabling asking of meaningful questions not possible through conventional a priori-hypothesis based studies.

The Ideas and the Concepts

Our view of the world has several different aspects, some overlapping and some not overlapping with each other. Science has been primarily driven by certain rigid assumptions about its methodologies. And science has fundamentally been reductionist in nature with the study of simplifications rather than actual phenomenon, striving to provide the simplest possible explanations. Einstein once asked “why the universe is comprehensible”? One answer to this question is that we, humans, with our rigid philosophy make it comprehensible, and then wonder about the complexity of the whole universe and make such statements.

Complexity is a term that should be used for anything that is incomprehensible and or undefined. At one pragmatic level, if at all, systems or processes or other conceptual aspects can be considered to be composed, and that too of parts, then there arises a serious problem in understanding systems or processes. Change is one fundamental aspect of any system or processes. Change gives rise to variations and thus to information. Complexity or the incomprehensibility or undefined nature of systems or processes should be treated, in one way, using concepts of the conventional rigid philosophy of science and also strongly breaking away from that rigidity. Now if complex systems are composed of innumerable parts and their interactions, then the cause effect relationships between the parts becomes obscure and rather becomes irrelevant. But according to the conventional philosophy of science, we understand the world in terms of cause and effects. Thus, in complex systems, one fundamental need is to study the systems or processes in terms of the variations and correlations within and among the various parts indicating the nature of complexity of such systems or processes. The dynamics of complex systems or processes in generalized phase spaces should be tracked irrespective of their cause or effect status. The holistic view of complex systems, in terms of its parts, can be got if we discard the cause-effect apriority hypothesis and determine the entities and their interactions, and their variations and correlations. This will give us the holistic structural and dynamical aspects of such systems/processes and give an understanding of their complexity. Now the irreducible complexity of mathematics inspires a theory of irreducible complexity of life, with some inherent information, noise and or randomness underlying the laws of life. To study complexity of nature and its irreducible characteristics of nature, the laws governing nature at various scales of abstractions and projections need to be found out and studies in comparison to each other to see the nature of the laws and structure of the grand unified theory of the world. Again, one obstacle in this path is the observations and interpretations of us humans which creates a relative reality in terms of us and makes it difficult to judge the absolute reality. It’s even not understood whether the reality we decipher is really relative it is actually the absolute. This is one ultimate level of complexity and incomprehensibility or undefined-ness. Even at this level, if the conventional philosophy of science is pursued, it will only give us more biased interpretations of reality and we will probably never be able to achieve our goal of understanding and deciphering the absolute reality. The irreducibility of complexity, if at all is there in nature, must be observable in some representative phenomenon of nature and life is one such setting which gives us a scope to investigate along those lines. One, due to the history of the nature of reductionist investigations performed on life and also due to the conventional rigid philosophy of science. The grand unified theories of our world, the conventional physics strives to decipher, e.g. quantum gravity, string theory are born out of reductionist nature of our thinking. Thus knowledge has evolved in a reductionist manner but since it's the knowledge about the complexity in nature, it is potentially a false view of nature. What should be pursued is the understanding of the structure of laws of different aspects of nature and then trying to merge those laws into a single law, which is primarily representative of the nature of the laws at the various levels or scales of abstractions and projections. The dimensionality of the laws, the dependencies of the entities and their nature, the fundamental parameters which can give rise to those dimensionality and the dependencies has to be studied to understand the nature as a whole which is more towards the relative absolute reality from our perspective. To find out the irreducible nature of life or nature itself, all laws at various levels of abstractions and projection has to be discovered. If there is data at various levels of abstractions and projections, then from such data ab initio zero hypothesis based analysis to determine the variations and the correlations in those data and parameters has to be determined. The magnitude and range of those variations will give the relative significance of those parameters in the systems. Similarly, the degree and type of correlations among the parameters will give the dependencies among the dimensions. This information together will give an insight into the number and nature of laws governing natural phenomenon. This will further elucidate the nature of complexity of such phenomenon, the different laws governing such phenomenon, their numbers and their dependencies etc. this will create and understanding of the fundamental nature of such phenomenon and give us a basis for studying such phenomenon. For example, recent research [Arenas et. al. Physics Reports, 469, 93-1153 (2008).] into complex networks, generalized models of complex systems in terms nodes and edges, have indicated my long standing conjecture about the nature of the world being fundamentally determined by the nature of connectivity (and not the exact connectivity) being the prime determinant of the dynamics of complex systems, rather than the nature of the elements in terms of their basic linear or nonlinear behavior, being probably true. In complex networks, the nodes showing the maximum variations should be treated differently from nodes showing minimum variations in some aspects. Nodes showing maximum variations have undergone changes and nodes which does not show variations has not undergone changes.

In the context of the living world, while studying life, one level of abstraction is the molecular level. This we are not considering as something arising from a reductionist point of view but rather from the view that if different levels of abstractions and projections are observable, then those levels should be studied, irrespective of which/ at which level actually they are, and establishing the variations and correlations will give us the complexity at that level, indicating the nature of the laws governing such levels of interactions. In life, there are dependencies across levels of abstractions and projections, which make the complexity of life a class apart from the complexity of the rest of physical nature. One part of this is due to the way we have studied biological systems or processes and other physical phenomenon, but there is some degree of inherent-ness of this quality in life or living systems.

With respect to the discussion above, the variations indicate the linearity or non-linearity within the parameter implying whether any linear or non-linear process is affecting the evolution of that parameter. The information content and computation capacity of systems or process are somewhat positively coupled with each other. The variations indicate the general symmetry within the complexity. Minimal variations indicate that either weak dynamical processes has affected such parameters, or otherwise strongly but nullifying dynamical process has influenced that parameter. Every parameter is equivalent to an operand, and the influencing phenomenon is the operator. Weak correlations indicate weak dependencies between variables or entities. Strong correlations indicate strong dependencies between entities or dimensions. Weak correlations also generate the possibility of multiple laws affecting the interactions between dimensions while strong dependencies raise the possibility of single or few laws governing the interaction between the dimensions.

While considering these discussions, in this context, it must also be stated that since, in lower organisms we do not differentiate within parts of the whole like mind and body, in higher organisms, we must conceptualize such higher organisms with distributed intelligence and also consider lower organisms with distributed intelligence. Thus the genetic networks also form part of the distributed intelligence. This is also relevant for the holistic approach to intelligence.

Various laws are triggered at different levels of abstractions, and this can be a very fine point of triggering which cannot always be determined. Observable resolutions of abstractions of phenomenon are important to determine the triggering of new laws and dimensions and interactions at different levels of abstractions. But at least that some law or dimension has been triggered can be inferred from lower resolutions as certain observables become strong observable at a certain level of abstracts making it allowable to differentiate between the triggering of new law/dimensions and absence of new law or dimension. When a systems or phenomenon evolves with an inherent randomness embedded in the process of evolution, then the final architecture, organization, dynamics of such systems or processes are not known before hand as randomness had strongly influenced its evolutions and from a large but finite set of solutions, which solutions has been generated is impossible to ascertain from abinitio apriori hypothesis based approach. Thus, the variations and correlations of the observables and their dependents have to be measured and the dimensions, dependencies and laws governing the systems can be deciphered from such an exercise. Otherwise the systems itself will remain unknown although it will be functional systems. This is a basic difference between artificially and naturally engineered systems esp. visible in life. If a matrix or vector be constructed of all the observables of a complex system which can be represented as a complex network, then the distribution and degree of correlations will elucidate the number of laws governing such systems and the degree of irreducibility of such systems. The objective is not to find out any cause effect relationships but to understand the systems as whole. The imposition of cause effect relationship and the act of their discovery impose s another level of reductionism on the system. That the assumption is that complex systems are composed of parts is of course one aspect of reductionism imposed on the systems. Any complex behavior or structure should be represent-able by linear systems, another form of reductionism. But this form of reductions may be considered for the better understanding of the holistic world. Thus any complex behaviors should be represent-able using simples of the behaviors and their combinations.

The lack of dependencies between the parameters is as important as the dependencies in the theory of complexity (as being propounded here) as this is important information in the holistic context of understanding the complexity of the systems under study and giving an understanding of the nature and number of laws governing such systems. The word system itself is controversial and difficult to define. But here it is being used in a very broad and fuzzy sense. The theory being propounded here is not directly dependent on the definition of the systems. So the attempt to define a system is not being attempted here. The uncertainties in the world and life are part of the complexity of world and life. And the degree of uncertainties and their contexts are necessary information for deciphering the nature of the laws affecting complex systems. The symmetry of laws, dimensions and their dependencies esp. the nature of these possibility observable entities is the super-symmetry in the world within the context of complexity.

While analyzing the cause of any phenomenon, the control experiment study theory will not be valid here because it's a complex network. Sometimes the effect of a certain entity will be observable at other times it will be not, this is because the entire phenomenon will be holistic. So excluding a particular parameter as a cause based on simple statistics will not hold here. The possibility of chaotic behavior also complicates the study of such systems, because at what resolution of change in signal will trigger a chaotic behavior is extremely difficult to determine or observe. With some data points, some basic and complexity-theoretic fundamental features can be inferred from some small number of measurements of a particular system if the measurements contain observation points suitably reflecting the features of the systems or process. Variations also indicate the degree of evolution in the sense of changes, like quantifying the degree of changes that has affected the system or processes, tracking the phase space trajectory of those changes to understand the basic steps through which those changes were brought about. The number of dimensions needed to describe a system or a process is an indication of its complexity. Some fundamental questions naturally arise from such an approach. At what level of abstractions new dimensions arise? How the dimensions interact to generate the fundamental complexity of systems and processes, at different levels of abstractions and projections? This involves the deciphering the state vector of biological systems and processes.

Complexity theoretic approach has the potential to uncover counter intuitive results and inferences, which are not observable with conventional reductionist and other approaches. It's a challenge for complexity theory to establish itself with our entire education and science being dominated by conventional reductionist approaches. It will be very difficult to even appreciate the value of the results obtained by complexity theoretic approaches if we continue to think in non-complexity theoretic process. Complexity theoretic approaches also help to uncover underlying unrecognized principles dependent upon both natural and artificial phenomenon, of even artificially engineered systems. Integration of diverse plausible inferences on complexity of systems is required to achieve a more holistic understanding of the complexity of the system, rather than its individual components. This is also leads from data to a holistic view of biological systems, in a new way. It also raise the possibility that complete data is not required for a holistic view and that this line of investigation is necessary to pursue to reach a holistic view. This is neither top down, not bottom up in the strictest sense but is more of a cyclic process between the two approaches with several other less-defined caveats. This is not multi-scale modeling also. This is a approach where the algorithm of inference is essentially fuzzy and case-sensitive but, in all strives to get an holistic understanding of the nature of biological systems from different levels of abstractions and projection, integrated together, not in a defined way but in a fuzzy way but reaching towards a general understanding of the nature of the system and phenomenon. Thus must be interpreted in the context that variations and correlations in a connected system gives the information content, degree of compression (into laws) and the degree of randomness and the indication of the independent set of laws or dimensions that have made the system. The more the correlation between the systems, the more there is a single law. Less the correlations between various parameters of the system, less chance there is a single law. More the variations for each parameter or less the strength of patterns, more number of laws or processes has affected the parameter. Of course several forces can balance out net effects, but this has to be dealt case by case basis. The ab initio discovery science method is suitable because it does not fundamentally assume any a priori hypothesis or law and thus can lead to the proper description of the actual laws of the system.

Evolution of organisms can be considered as multi-objective optimization process. The environment of the organisms is the multi-objective function that has to be satisfied by the evolving systems i.e. the organisms. The evolution process takes components available in the environment i.e. the biosphere and synthesizes the circuits or systems i.e. the organisms according to a process that involves process that have random aspects. Since the process of evolution is so complex and since it has been not been observed or is not observable in its entirety in a defined/deductive manner, due to its complexity e.g. the components, their interactions and the state values of the components and the interactions ths the evolved circuits or systems’ complexity is unknown and uncharacterized. It is of utmost importance to characterize such a system to enable an/the first understanding of the evolved solution and its degree of complexity although measures of complexity are still not well-founded esp. in terms of biological systems which have undergone a complex adaptive process of evolution. The interacting components and their interactions, their type and number and the state values they take in the systems must be delineated to enable the first understanding of those systems in reference to this viewpoint. In general it is believed that biological systems or organisms have evolved o be composed of innumerable components and types and a variety of interactions and types. These components and interactions take state values that can be classified as continuous, semi-continuous, discrete, sorted enumerative and unsortable enumerative. In contrast to a priori hypothesis/fact based deductive reasoning, these systems and processes or phenomenon demand an ab initio no apriori hypothesis driven inductive analysis.

The partial information regarding the complexity of these systems e.g. the organisms or major subsystems within the organisms (e.g. the signal transduction system) is distributed in databases in sometimes structured but mostly unstructured form. Thus to delineate and characterize the complexity of such evolved adapted systems which as undergone a semi random process of evolution, the need is to develop computational methods which may be an example method to collate information from distributed sources and put together to make the analysis of complexity amenable. Thereafter, the need is to produce a description of the complexity of the systems or subsystems e.g. signal transduction in taxons to get the first ab initio non-hypothesis driven picture of the nature of the systems in terms of their components, their interactions or dependencies and the state values they take in the present manufactured systems i.e. extant species or taxons. This would give the initial picture of the internal circuitry of the systems and in further on would enable to grade or pinpoint components and interaction and state values for their role in the structure and dynamics of the systems or subsystems. This path from description of the complexity to the grading of the parameters for their relative importance is an undefined and most probably a colossally complex process which would require multiple or may be innumerable steps of analysis. This is primarily due to the unobservability of the complete process of evolution and the multi-objective function due to the randomization component involved in the process and environment and the large variety or complexity of the process and the environment from which the process of evolution draws resources to construct the solutions. Thus the evolved solutions are of completely unknown complexity in their internal architecture which can be very less or very high. Since the complexity of the environment or biosphere which other acts as the multi-objective function that needs to be satisfied and also resource source for the evolution process is unobservable in its entirety, an since the steps in process of evolution are also unobservable in their entirety, so the manufactured systems system demands a complexity theoretic (here ab intio no-hypothesis driven characterization) analysis.

Deductive a priori hypothesis driven analysis seems to be playing with trial and error to understand such colossally complex adaptive evolved systems as they may or may not enable their solutions or understanding. Whether they will be understood or solved through a priori hypothesis driven deductive reasoning depends on the actual complexity achieved by the system. If they are simple, then this way of reasoning will be successful in understanding them. But whether they are simple or more complex or colossally complex cannot be understood before hand. So it comes to an issue of choice of which of reasoning our analysis path to take to understand such systems. Ab intio non hypothesis driven complexity theoretic analysis is another way to analyze such systems. But the main limitation of this approach is the unobservability of the process of evolution, the multi-objective function and the biosphere as the source tank in their entirety which limits a full and complete complexity theoretic analysis. But classifications of systems, process and phenomenon at the upper levels of hierarchy of the classification trees provide some suitable abstractions and projection or reference frames where such complexity theoretic yet somewhat (but no grossly) partial analysis can be performed. Thus here selective the levels of abstraction of 3 kingdoms of life and the fundamental functionality of living systems e.g. control, communication, metabolism enables a semi-complete but an acceptable level of complexity theoretic analysis. The fundamental challenge to such analysis is the availability of the amount of data which is again related to the observability of such systems. At present since large amount but grossly inadequate data is available, the challenge comes to the computability of the integration of such data in such a manner to enable the complexity characterization of such a systems or phenomenon e.g. signal transduction subsystems in the major kingdom of archaea and the comparison and unique and common parts of archaea with the remaining 2 kingdoms of live. Such analysis demand a level of abstraction that is high enough to form a complete set i.e. the 3 domains of life or fundamental functional subsystems/sub processes of living organisms. But the limits of computation also come into play when the analysis is at such a high level of abstraction involving innumerable sub components and sub interactions. Thus it’s extremely difficult to choose a suitable path of analysis but yet this must be done. (This paper is an attempt in such a direction.)

The study of the complexity of life i.e. “BioComplexity” demands completely new approaches to view Life with considerations of mesoscopic connectionist integration of projections, abstractions, dimensions, interactions, holism and individualism, among others. Integration of different viewpoints to study BioComplexity of “(bio)system(s)” demands a deviation from inference-based cause-effect analysis to ab initio characterization of the complexity of the entity (/event) with no a priori hypothesis. Modern BioComplexity research needs exemplifications of the above through (A) formulating suitable problems, (B) demonstrating approaches to their solutions and (C) studying the outcome of the solutions. One aspect responsible for BioComplexity in the living kingdom is the signal transduction system which a suitable substrate for such a problem because it incorporates several projections and abstractions of the dimensions and interactions from the genome to the organism level in generating the unity and diversity of the living kingdom. It must be attempted to 1) demonstrate an approach to a suitable mesoscopic integration to study signal transduction and 2) its subsequent application for ab intio global characterization of the signal transduction system in a important taxon of the living kingdom. One successful approach to such a problem, through computation, through a computational pipeline/protocol which is one suitable (mesoscopic i.e. from protein length to species distribution) integration of highly heterogeneous and multidimensional processes and information where the utility and complexity of the entire computational infrastructure is demonstrated. There are limitations to this kind of approaches which have to be discovered & listed by the scientific community. Considerering the above example, the global a priori/ab initio view of the signal transduction proteins (families) of the archaeal kingdom and their properties (e.g. length, domains, species) within & across themselves and their sharing & comparisons with the eukaryote and bacterial kingdoms has to be delineated. This kind of complexity theoretic approach, for life, provides an understanding of the general characteristics of the signal transduction proteins in the archaea (or similar problems) and more importantly naturally generates some important line of investigations. The archaeal kingdom is arguably the least understood among the three kingdoms of life. The archaeal kingdom shows properties of life that were previously not thought of and is extremely difficult to explain. The process of biological evolution which has been responsible for the creation, maintenance and propagation of the archaeal kingdom needs to be understood on its own and also in comparison to the phenomenon in the other two major kingdoms of life. Work should attempt an exercise in the delineation or characterization of the complexity of the signal transduction network components or proteins in the archaeal kingdom. Also, works should attempt to put forward a philosophy of ab initio characterization of complexity of colossally complex systems like life forms and demonstrates a computational procedure to tackle such complexity using available yet incomplete data through large scale data integration and data mining and data visualization and software integration. Several established data bases and data retrieval systems must be put to use all together in an integrated framework and linear and nonlinear (clustering) data mining protocols to be used to extract the complexity patterns in the biological systems/entities/phenomenon under study. The computational processes may well be a repeatable pipeline with wide applicability. This wide applicability also has to be purposefully bounded in its boundaries (upper and lower) to approach problems which are not too general or too specific which is the problem of current biological computational infrastructure. One of the major difficulties that is to be faced during the development of this kind of computational pipelines/ methodologies is the frequent and unpredictable changes and/or updates in the database formats and design and accessibility. This prevents the smooth access and processing of data in the pipeline as and when the changes occur in the databases. This is a natural and unavoidable aspect of the ever dynamic and perpetually evolving nature of biological databases whose complete domain and ranges can never be defined in full.

The integration of (such) large scale information of complex systems requires computational methodologies that should reflect the complexity of the systems under study. The design of engineering systems e.g. the computational methodologies, should mirror the natural systems i.e. their complexity to effectively observed and control such nature systems. This also makes the engineered systems robust to the uncertainties generated by the natural systems. While studying complexity, this aspect arises and the design of systems for observing, analyzing and controlling complex systems becomes manageable if such concepts are implemented.

Because of evolution and a long history of environmental accidents that have driven processes of natural selection, biological systems can be regarded as engineered entities than as objects whose existence might be predicted on the basis of the first principles of physics, although the evolutionary context means that an artifact is never “finished” and rather has to be evaluated on a continuous basis. A key aspect of biological complexity is the role of chance. One of the most salient instances of chance in biology is evolution, in which chance events affect the fidelity of genetic transmission from one generation to the next. Perhaps one of the most striking manifestation is that individual biological organisms—even as closely related as sibling cells—are unlikely to be identical because of stochastic events from environmental input to thermal noise that affect molecular-level processes. Complexity, then, appears to be an essential aspect of biological phenomena. Thus modern biology needs to integrate a number of diverse intellectual themes e.g. that of the reductionist and systems approaches. Every living thing is unique. Furthermore, the uniqueness and historical contingency of life means that for population-scale problems, the potential state space that the population actually inhabits is huge. The study of such unique entities or phenomenon demands the integration of computing into biological research. But this integration of computing into biological research raises epistemological questions about the nature of biology itself. The question of how biological phenomena can be understood has not been fully settled.

Every system or phenomenon undergoes changes in its design or dynamics, either naturally or artificially with time or in generalized phase space appropriate for the context. According to one of the most accepted view of evolution, the process of evolution is a kind of multi-objective optimization process where inherent random noise percolates through the biosphere and influences the trajectory, to a significant extent, in the phase space of solutions of stable yet open systems i.e. individual organisms. The role the entire biosphere plays in this complex evolutionary process is observable through the constraints set by the bounded yet changing environment and the network dynamics of the organisms within and across themselves to create the ecosystem, which, in turn, affects the process of evolution- a cyclic and recursive phenomenon. Although evolution is partially synonym with the history of life on earth, yet it must be taken into consideration that the present is not just a continuation of the past, but rather emergent from the past. The mix and match process of biological evolution traversing a complex multidimensional phase space in search of optimal or minimally optimal solutions inherently destroys the conventional statistical nature of the biosphere at various levels of abstractions and triggers the emergence of network behaviors. These networks can respond to external or internal perturbations i.e. either information or noise, as a whole and maintain the dynamic stability (robustness) of the entities e.g. organisms. Yet these networks allow the organisms to execute the inherent property of adaptability within the complex uncertainty of the environmental constraints or demands.

The colossal complexity of biological systems or processes, much more complex than most other physical phenomenon necessities various approaches to the understanding of this complexity. Ab-initio no apriori hypothesis based holistic and global characterization and classification of biological systems or processes at specific levels of abstractions and projections is one such approach proposed in this chapter. This is esp. true for biological systems or processes which are poorly understood or which are thought to be a prime contributor to the total complexity of biological systems. With the availability of large scale data at different levels of abstractions of biological systems esp. molecular i.e. DNA and Protein sequence primary and secondary data, and the availability of high-end computational infrastructure on desktops, it is necessary design, develop and implement computational processes or systems e.g. computational pipeline or protocol to target general and specific areas of biological problems. It must again be mentioned that current computational systems or processes either focus on the most generalized aspects of biological systems or extremely specific aspects of such systems or processes. But with the affordable availability of data and computation power, it is necessary to develop computational systems or processes which neither are to general nor too specific to any particular aspect of biology but which can solve problems of intermediate nature without complete automation but with intermediate human intervention and appropriate scalability according to the requirements. Such computational systems or processes are not the focus of current bioinformatics which mostly fluctuates either completely passive software or completely automated software.

Thus, in one way, based on all our previous arguments, we can approach the problem of biological complexity from the level of abstraction of proteins and protein families for the signal transduction systems which is a prime contributor to the complexity of biological systems in the archaeal kingdom of life which is the poorest understood among all the three fundamental domains of life. We can approach this problem by developing a computational pipeline or protocol which uses data mining and database integration and software integration using a specifically selected global probabilistic graph/network clustering algorithm that successfully classifies all the signal transduction proteins in the living kingdom into protein families. Subsequently the computational pipeline can be used to delineate the numerical distribution of signal transduction protein families in various taxons combinations. Thereafter, the pipeline can be used to characterize the complexity in terms of the variations and correlations in the signal transduction protein families in archaea wrt the numerical parameters as protein family size, protein length distribution, taxonomic spread of the protein families, copy number distribution in the containing species, domain composition and their frequency distribution, domain architecture variety and distribution, maximally used domains in the families etc. The above characterization can be done for signal transduction protein families which are specific to archaea, and other signal transduction families which are present in archaea and also in the bacterial and eukaryotic kingdom.

The first and foremost requirement of studying a complex system for whom large-scale data for various parameter or state variables are available is the clustering and thus classification of the data set to identify what component groupings are present and what are not based on various relationship criteria among the parameters or variables, and subsequently characterize their nature or their complexity i.e. their variations and dependencies within and across themselves to better understand their existence and interactions among them that make up the whole complex system for which the analysis is being done.

Clustering or grouping of genes and proteins based on any parameter(s) that characterize each of the members of a set of genes or proteins enables to understand the basic partitioning of the set and helps to understand the fundamental causes of variability of the data set of genes or proteins. Clustering can be used to discover the families of genes and proteins. For example, if a dataset contains the sequences e.g. amino acid sequences of proteins, then certain clustering techniques can be used to find out how the set of sequences can be partitioned into groups based on pair-wise sequence similarity. These groups can be thought to correspond to gene or protein families. This problem ideally requires an algorithm that considers the entire similarity network of all the component of the original set as a whole and discovers the more accurate grouping within the set. Also, to analyze the groupings generated by such algorithms, there is a lack of software (as discussed previously) to uncover the meaning of the groupings/clusters based on ancillary information (information besides the sequence data). This is required to analyze the biological meaning of the groupings and the distributions of the data. Thus there is need for software that would automatically annotate the groupings/clusters of the proteins with relevant biological data.

The gene/protein families can be considered as the basis set of functions present in an organism whose components [like vector components] interact with each other to produce a functional dynamic network at the molecular level. The macroscopic evolution of organisms can also be visualized as the evolution of the basis set i.e. the gene/protein families and the interactions among their vector components. The same is true of taxons which groups of organisms related by phylogenetic lineage. The study of the allocation of the subsets of the full basis set (of the entire living kingdom) in sets of organisms i.e. taxons is an important way to understand the similarities and differences among the taxonomic sets in terms of basis set of molecular functions.

The living kingdom at all scales of abstractions and projections is extremely complex. One way of interpreting the phrase of “extremely complex” is to consider the living kingdom, at any level, a hugely complex network made up from innumerable nodes and interconnections and an equally diverse types among the nodes (entities, parameters, elements etc) and their interconnections (relationships, interactions etc) (where sometimes the interconnections are nodes and the nodes are interconnections). The scenario is further complicated by the “unity and diversity” in the living kingdom where every entity has its own uniqueness that necessitates its study as a separate individual and also its commonality with other entities that necessities a gradual generalization of similar types of entities. This intriguing “unity in diversity”, quite different from other physical systems, apparently eludes biological systems from sweeping generalizations, possible in other fields of science. The complexity in the living kingdom may be partitioned between “trends” and “events” where trends denote the common patterns found in a set of entities, and event denote the structure and dynamics of individual entities. At a level, these two tend to merge and blur the distinction between. Thus in the context of the living kingdom, by complex, we mean the enormous diversity yet similarity in the variations and correlations among the nodes and their interconnections. But the sequencing of the genomes provided a new approach to biological sciences- to delineate the complexity ab initio of living systems without any a priori hypothesis. This approach is a subtle one where the science moves beyond the usual cause-effect study or control experiment study. This approach opened up one of the possible ways to address the complexity of the living kingdom- that when the system is extremely complex, it can be studied by delineating all the nodes and their interconnections and subsequently studying the variations and correlations of the nodes and interconnections and thus establishing the network of the living kingdom. This would subsequently enable the more conventional scientific study of systems w.r.t. e.g. the cause-effect relationships etc. To be a little more specific, since it is accepted that interactions exist in the living kingdom between several different levels of abstractions e.g. environmental temperature (environmental level) can influence the dynamics of gene expression (molecular level) which affects the behavior of the organism (organism level) which affects the social behavior of the population (population level), an alternative approach to understand the living kingdom where the entities responsible for this complexity is first identified and their dependencies (along with the strength of the dependencies but not trying to identify the independent and dependent variable because in a colossally complex network, it is not wise to pinpoint early which is the independent and which is the dependent variable) are identified is always desirable. This would create the MAP of the living kingdom at desired levels of abstractions and projections. The need is to create such maps and merge them when possible to get a suitable picture of complexity of the living kingdom. In the same spirit of large scale genome sequencing, not every individual sub or component map may provide conclusions, but every map goes ahead with the delineation of the complexity of the system that is step forward in the deeper understanding of the system with the addition of subsequent maps.

Biological Complexity can be considered to be composed of several reference frames with projections across each other and composing different levels of abstractions. These reference frames are further composed of several dimensions not necessarily orthogonal or orthonormal. Every biological concept or entity is observable or measurable with respect to any of these several reference frames. E.g. A set of reference frames are the levels of abstractions ranging from molecules to pathways to cell to organisms to populations to ecology. Another set of generalized reference frames are the structural and dynamical aspects within and across the above reference frames. The complexity of this biological entirety also has a soft fractal type nature in that several reference frames can be considered to compose a set of other reference frames. The state variables that define biological complexity in these reference frames have another level of complexity. They have several dimensions themselves. It may be hypothesized that the dimensions that make up the reference frames are not themselves basis sets, but rather a composite of other dimensions. The nature of biological complexity is such that these composite dimensions cannot be resolved into orthogonal or ortho-normal dimensions in any single refernce frames.

Signaling and communication within and across cells is a prime aspect that determines the biological complexity. Signal transduction is a fundamental biological function that enables an organism to respond to stimuli and communicate with its environment, enables growth and development. Signal transduction also forms a part of the regulatory machinery of the cellular processes. The diversity in the structure and dynamics (across time scales) of signaling systems strongly influences the diversity within unity across the living kingdom. The concept of viewing signaling systems from the perspective of gene families is one of the reference frames of measuring biological complexity. Further, taking the projections of this reference frame across another contextually relevant reference frame, that of the tree of life, will elucidate the complexity of the signaling sub-system within and across taxons. Grouping organisms into evolutionarily related groups or taxons creates another level of abstraction that acts as a super reference frame, super with respect to the organisms level reference frame. Also, as stated earlier, the gene families state variables of signaling systems can be described in terms of their component parameters that again are nether completely orthogonal or ortho-normal i.e. parameters such as biological functionality of the family as a whole, length ranges of distributions of the members of the protein families, number of different species in which this family is present for a particular taxon or kingdom, their number in the entire taxon and also their copy numbers in the organisms composing the taxon. In addition to these, the number of different domain architectures making up the members of the family, the number of modules making up the architectures, how much each architecture is predominant in the whole family and what is the distribution of predominancy, no of unique domains/modules making up the architectures, frequency distribution of the domains in the whole family and the most abundant domains in the whole family. Of course, it is never claimed that these parameters compose the state vector of the protein families completely (as also easily deducible from the opening arguments of the nature of biological complexity, its reference frames, its composing dimensions and the nature of the biological state vectors defined in those dimensions), but from a certain frame of reference they do give a complete picture of the complexity of the protein families of a particular generalized biological functionality. This certain reference frame can be considered to be the structural nature of the members of the families and their basic tendency for abundancy in systems in which they are present.

With the unstructured and strongly non-reducible (weakly reducible) nature of biological complexity, as also supported by our opening conjecture on biological complexity, it is imperative that sets of biological dimensions be chosen for study to reveal the underlying complexity of biological systems in the approach of exploratory discovery science. Thus, with the availability of large-scale data at various levels of abstractions and projections of biological systems, it is important to take up exercises to define contextually significant state variables and (automatically/computationally) organize that available data to bring out the net patterns i.e. fundamentally variations and correlations within and across that data in terms of the defined state vectors, to reveal the underlying complexity of the systems in simple terms. The nature of biological complexity, in spite of being strongly irreducible, will be elucidated gradually by the above process of investigation.

An engineered machine can be divided intro three fundamental parts: 1) the central or core or the engine, 2) the controlling parts and 3) the communication between the engine and the controlling parts. The engine is the main driver of the machine. Its structure and dynamics determine the identity and fundamental behavior of the machine. The controlling parts regulate and fine tune the functioning of the machine and adapt its behavior, dynamics to external requirements or conditions. The communication systems establishes the necessary message passing and signaling between the controlling parts and also between the engine and the controlling parts for accurate and efficient transfer of signals from one part to another. Of course all the three parts are equally important to the functioning and longevity of the machine, but the communication system in intricately coupled with both the engine and the controllers for proper transduction of signals across the entire machine and its interactions with its environment. In biological machines, which are naturally evolved and not human-engineered, the communication system is inherently and implicitly embedded throughout the machine at different levels of abstractions and projections. The colossal complexity of living systems is to a significant extent is manifested through this signaling and communication network within and across living systems and with their environment. But amazingly, this communication system itself it extremely complex, as are several other biological subsystems and is one of primary substrate for the evolution of the “unity within diversity” in the living kingdom. The fundamental components of the signaling systems, at the molecular level are the signal transduction proteins (and their protein families). The similarities and diversities within proteins and of course the similarities and diversities of the above similarities and diversities across different organisms or taxons is a prime contributor to the similarities and diversities within and across the living kingdom, esp. in morphology and behavior across different scales and abstractions. This necessitates the study and understanding of signaling and communication system's complexity in the approach of identifying and delineating the fundamental components of the entire systems and sets of dependencies within those system components. The prime indicators of the complexity of the signaling system or in general any molecular level system is the biological role of the protein families constituting the system, the size of the families i.e, the number of members indicating the importance of the family in the organisms, the length ranges and distributions of the protein members- an indication of the information capacity and functional complexity of the members and relative diversity in terms of the informational complexity among themselves i.e. same functional class, the number of species in which they are present- indicating their spread across the living kingdom and thus their generality or specificity of utility within various organisms or taxons, the copy number of the families in the organisms in which they are present indicating their tendency for duplication within their host genomes and thus their volumes of requirement within the organisms and capacity to evolve further. Also other parameters like the no of different domain architectures making up the family indicating the complexity of the families evolutionary history and its tendency to diversify in structure and function, the length of the domain architectures and their distribution among members indicating the degree of modularity of the family members and relative distribution among equi-functional set, the frequency distributions of the domain architectures within the families indicating the presence of any preference for any architecture among the family members and thus any bias towards a particular architecture or form of modularity, the no of unique domains making up the architectures of the family indicating the potential or degree of multifunctionality among the members and thus their functional complexity, a prime factor in signaling and communication systems, the frequency distribution of the unique domains among the family members indicating any bias for any or some particular domains and thus the strongest basis for their homology/common ancestry and functionality, and last but not the least identification of the domains which are present in maximum number in the family, thus giving a strong indication to their major functional role in the entire signaling and communication network. The actual values of the above parameters and dependencies or relationships within and across these parameters, within and across living kingdoms will provide a complexity map and an extremely valuable insight into the colossal complexity of the signaling and communication system of the living kingdom.

The archaeal kingdom was identified as a separate kingdom of life at the last, long after eukaryotes and bacteria, and after that to the worlds amazement, it was being slowly discovered that the new kingdom is uniquely different from the other two kingdoms. With the generation of large-scale genome data, more fascinating facts are being discovered about the very elusive archaeal kingdom, but there is a long long way to go before we can understand this new kingdom tp a minimum level of satisfaction. This extraordinary and intriguing uniqueness of the archaea immediately compels us to focus on its signaling system, to see what kind it is, how much it is similar to that of the other two kingdoms of bacteria and eukaryotes and how much it is different from them, how much the dependencies are similar and different within and across the kingdom or lineage specific protein families, within and across kingdoms and lineages. This immediately demands the creation of a complexity map of the archaeal signaling components i.e. i.e. the identification and delineation of the variations and dependencies or correlations within and across the archaeal signal transduction protein families in terms of the fundamental parameters determining its complexity i.e. their biological names and functions, family sizes, information capacity and functional complexity (lengths of protein members), their taxonomic spread and copy numbers, their variety of domain architectures and their length and frequency distribution among the members, their modularities and multifunctionality i.e. component domains and their usage and frequency distribution. This necessarily demands a large scale at least semi automatic computational analysis pipeline for addressing this problem.

Any systems have architecture and its dynamics. This is a reflection of its complexity. Both the architecture and the dynamics are transformations of the information content and computational capability of that system. Any system is created through a process and thus reaches a particular form and is able execute a process. How (type) the system is created is the algorithm for its creation and what (type) the system can do is the algorithm it can execute and transformations of its architecture and dynamics is the information content of the system, which contains the information about the system per se, its algorithm of creation and the algorithm it can execute. Biological systems have not been viewed in this manner specifically to find the validity of the theories of information and computation and communication across the levels of abstractions of projections of these systems. Due to the availability of adequate data at the various levels of organization of biological systems, this theory can be developed and its validity can be tested in a dynamic iterative process with continuous analysis of biological data for the various variations and correlations, similarities and diversities, complexity and evolution, organization and dynamics, but with an ab intio view of none-aprori (conventional) biological hypotheses to decipher how biological systems are information systems and computational systems and is their engineering design, their evolution and their functionality in the sense of theories of information, computation and communication. To my knowledge such an exercise to seed the creation of such a field of investigation has not been initiated which revolves around the cyclic iterative dynamic process, between theory and data, of construction the theory of biological systems as pure information, computation and communication systems. I feel deep insights into new interpretations of life will be gained through these form of exercise as embodied in this paper and will hopefully be pursued by other people who realizes the significance of this exercise. But it must be appreciate that discrete and defined conclusions should not be searched for as by the philosophy of science, control information will be available only after several exercises of similar kind, which will create a platform of scientific inference from a population of possible incomplete inferences. Also, the distinction between conventional biological hypotheses based approach to understand life must be distinguished clearly from this theory primarily wrt to the final objective of this form of exercise which is to understand biological systems as true information, computational and communication systems. As an example, individual gene families or gene groups can be studied wrt to entities (natural as well as defined/abstract) at several levels of abstractions (where adequate data is available) e.g. nucleotide, dinucleotide, codon, nmers, amino acid, domains wrt to different parameters e.g. (gene count, synteny, codon choice and usage, domain distributions, gene & protein lengths) to study their complexity and evolution across species to characterize the organization of biological information within and across kingdoms of life, within and across architectural and dynamical classes, and get preliminary idea (as this form of investigation is nascent) of the computational capacity of that biological information and also its evolutionary algorithm for its creation.

From all the arguments and discussions above, it can be summarized that the genome of an organism can be considered as an information system encoding the functionality of the organism, either completely or partially. The genes at one level of abstraction can be considered the units of functionality. Gene families or protein families are one set of fundamental units of functional or basis sets that are functional dimensions of organisms or groups of species i.e. taxons. But the most fundamental unit of functional or basis set of dimensions is the domains which combine amongst themselves to create functionalities in general, and generate innovation in functionalities. The organism or species or taxons are the units of systems which are composed of these unitary dimensions. The complexity of these functionalities and these system units or taxons, in one way, are best studied by the complexity of the functional dimensions composing them, the corresponding weightage of these dimensions in the taxons, which would give us the state vector of these dimensions by themselves and the systems units which they compose. The symmetries and complexities, the conservation and divergence of the dimensions themselves and the system units they compose gives the complexity, I one way of the dimensions themselves and the systems units they compose. The genomes of organisms, species or taxons or the system units encode this complexity. The study of the complexity will enable us to comprehend the complexity of biological systems, their specific abstractions and the dimensions, both fundamental and higher, that compose these abstractions. The process of evolution, through several forms of mutation and selection, as a multi-objective optimization process, has generated the solutions to the dynamically changing environment or solution space and has continually adapted the systems to give adequate solutions i.e. organisms, protein families. The multitude of adequate solutions generated by evolution is composed of various dimensions and has their corresponding state vectors. The state vectors characterize the solutions. Not, deciphering the state vector, or more fundamentally their symmetry and complexity, their conservation and divergence gives an understanding of the complexity of evolutionary process itself and also fundamental design complexity of the solutions generated or created, in terms of functionality of the solutions (as for this paper work). By the basic nature of life and evolution that has created it, there is an inherent randomness in different levels of abstractions and projections of these, and a pervasive uncertainty due to the lack of information, and complete and closed generalized theory of life and evolution, which makes the observables unstructured (i.e. there are innumerable inconsistencies in biological data). This creates a need for designing systems that cal capture the complexity and also the uncertainty of life and evolution from specific levels of abstractions and projection to delineate the complexity of those levels. The fundamental process of human activity of classification is by itself extremely a hard problem in biological systems, because of the inherent randomness and non-linearities and coupling of the entire phenomenon of life, evolution and environment. Thus to classify biological data, esp. the fundamental functional dimensionalities of biological system is an open challenge problem by itself, but which, even in its crudest forms will lead to a broad categorization of the complexity of the evolution and life itself. Signaling, between systems and their environment is the prime factor in the adaptive process of evolution of these living solutions to the dynamically changing environment, since signaling is the mechanisms by which the solutions interact with their environment or objective function, within the solutions themselves as other solutions are also composing the environment or total objective function (the environment being composed of a part by itself and of solutions to the objective function, where each of the solutions influence both the inherent component of the environment also the other solutions which form part of the total environment or total objective function for a particular solution. It is also evident that the particular solution which is adapting to the environment is also influencing the environment.), and also each solution maintains itself with the process of internal signaling. Each protein family being a fundamental functional dimension, it is imperative to understand the fundamentality of the dimension. Since the proteins of each family can be conceived a symbolic sequence (of a finite alphabet sets- both the closed (amino acids) and open (domains)) encoding the functional information of the dimension, basically at two levels of abstraction, one at the level of amino acids, which is thought to be completely determined and thus a close and finite alphabet set, and the other of fundamental functional units domains – open and finite since not all domains have not been determined. When such dimensions are composed, then it is imperative that the internal complexity of those dimensions be understood, which creates the identity of those dimensions. But it must be considered here that there are inherent random components to this information. Thus an order of the information content is important to understand the complexity of those dimensions from different abstractions. The length of a protein sequence is, in one way, the measure of the information encoded. Thus for a protein family, by itself, and when viewed wrt to unit systems like taxons, the length range and nature of variations within that range is a measure of the order of informational complexity and the process of evolution that generated that functional dimensions from the level of abstractions of the finite and close alphabet set. The distribution of the functional dimensions, that is the weightage of the functional dimensions in each species/taxons gives the degree of importance of the role of that dimensions, whether it is a major factor or acts as weakly interacting factor in the systems units, is elucidated. The generality of the functional dimension within a set of species, given by the species spread gives the role of the strongly or weakly interacting dimension in the related system units, and indicates whether the dimension is specific to certain solutions or is general for a whole class of solutions. The number of domain architectures making up a protein family i.e. a functional dimension and their weightage of distributions within the family gives the idea how well-bounded that functional dimension has been in the process of evolution implying the stability of the dimension and its nature of evolution. Similarly the number of domains or unitary functional units composing those dimensions gives the complexity of those dimensions from the levels of abstraction of the unitary functionalities, and is a different complexity from that available from the length of amino acid sequence. It indicates the multidimensionality, multi-modularity and multi-functionality of the holistic dimensions of protein families. The frequency distribution of the unique unitary functionals gives the relative weightage of those dimensions in the holistic dimension. Thus, in totality, this perspective creates the architecture of complexity of biological functional dimensions, their internal inherent reductionist complexity, and the complexity they create holistically in living systems/units i.e. organisms, species or taxons. When this perspective is considered for the biological functional dimension of signaling, which is the dimension responsible for the interactions of life, evolution an environment or in other word responsible for the interaction and thus stability of solutions with other solutions and the components of the total objective functions, and thus influences the process of evolution of the solutions themselves (as a feed back mechanism), its create s a fantastically fascinating picture of complexity of the living world in general. This perspective becomes even more remarkable when such complexity is elucidated independently for the solutions or taxons which form a uniquely different class of solutions, extremely different when compared to the total set of solutions, i.e. archaeal kingdom and viewed in terms of the solutions themselves and wrt to the mainstream class of solutions. These perspectives would give us the generalized complexity of the biological solutions and the functional dimensions and also of the major exceptions of the living world. This perspective demands new kinds of computational solutions or protocols which can capture both the complexity, generalized trends and their major exceptions, and also the incomplete and open informational structure of biological information, such that the elucidation of the fundamental complexity of biological dimensions and solutions delineates both the understanding and its gaps of such an exercise.

Based on the discussions till now, below, I forward and document some of my own conjectures elucidating the ideas of complexity and biocomplexity embodied in this paper.

A plausible conjecture of the nature of totality is the complete connectionist nature of totality. By totality we mean the ultimate reality, including physical and antiphysical, the the mental and conceptual, the mathematical and the non-mathematical, and the linguistics aspects of our world, as perceived from within the boundaries of our consciousness. One fundamental aspect of totality is its representation and perception. How totality or any aspect of it can be represented and /or perceived? The question is whether there is a universal and unified representation or perception of totality or any of its aspects? Can such representation or perception be consistent with our historical views of reality or totality in general? Can physics, biology, mathematics, philosophy be represented in such a representation? But, more fundamentally, whether such a representation can exist at all? And also, how much general can that representation be made, without being inconsistent with our view of the world? Can such a representation be all-inclusive, in the sense that the complete knowledge of our world can be unambiguously represented in that representation? In this context I have a fundamental conjecture and some other fundamental conjectures that seem to be consistent with each other, complete, universal and unified at the same time.

Assuming that such a unified, universal representation of totality exists, I go on to conjecture that such representation is a connectionist one. My conjecture is that, any aspect, part or component of reality rather totality, has atleast a connectionist representation. The second and more fundamental conjecture is that the one and only representation of totality or any aspect or part of it is a connectionist one. By connectionist, I mean a network with nodes and edges, where nodes represent entities and edges represent relations (or interactions). This, I claim, is a universal and unified representation of totality, including the physical, mental, mathematical and linguistic aspects of it. If one sees through both the above conjectures, the first ensures that the second one is true- because, if any aspect has atleast one connectionist representation, then all aspects have at least one connectionist representation. If all aspects have at least one connectionist representation, then the only representation of all aspects is the connectionist one. Since the only representation of all aspects is the connectionist one, therefore, the one and only representation of totality is the connectionist one. Thus, the unifying conjecture is the second one- **“The only representation of Totality or any aspect of it is the connectionist one”. This I call the “complete connectionist conjecture”**. I call it complete because it is universal and unified. It can be perceived that this conjecture consistently includes the inherent interdependency of space-time, too.

This conjecture leads to another conjecture about the nature i.e. complexity of totality. The conjecture is “the irreducible complexity of totality”. The only representation of totality being the connectionist one, it is impossible to reduce the nature of totality to a single unified law. This is because of the connections distributed throughout totality. This defies the cause effect relationship of reality. One node influences the other and the other influences the previous one, directly or indirectly, making it impossible to identify the dependent and the independent entities. Rather, the concept of independence becomes weak in this context. This must be understood to be not due to duality of totality, as exemplified by logical positivism and logical negativism, which states that creation or existence of one aspect inherently induces the creation or existence of the dual aspect, of any part or sub-space of totality. Ofourse, the duality has its origins in conceivability and observability, computability and controllability, which are responsible for the validation of the creation of existence of entities of totality. This irreducible complexity of totality is purely due to the connectionist nature of it. You cannot reduce a network. This is the strong conjecture of irreducible complexity of totality. Later on, a weak conjecture of totality will be discussed which essentially is the irreducible complexity of totality in finite metric spaces, having origins in the non-linearity and perpetuality i.e. essentially again in the connectionist nature of totality. This strong conjecture of irreducible complexity of totality immediately points to the "inherent randomness of totality". If anything cannot be completely reduced, then we assume the presence of embedded randomness, usu. compensating for the lack of information regarding that thing. But, if totality is strongly irreducible, then it must be that totality is inherently random. Now questions can be raised on the distribution of randomness within totality. I shall not delve into that in this paper but will be dealt subsequently in another work.

The inherent non-linearity drives the perpetual evolution of totality. Thus, new nodes are created within totality due to the interactions of other nodes. This conjecture is on the "infinite recursive evolution of information". This essentially states that totality is open. New nodes, are created during the evolution of totality. This has to be perceived in the way that all nodes which can be "created" are present in totality, but totally isolated, with no connections with any other node. But its connection with atleast one node brings it into existence, and its subsequent evolution through the establishment of connections with other nodes, in parallel to the evolution of totality. This leads to the perpetual evolution of totality under linear (operating on neighborhood sub-spaces) and non-linear (operating on distant sub-spaces) processes. Non-linear process establishing connections between distant (in a connectionist sense) nodes leads to the phenomenon of singularity, criticality and emergence, esp. through the creation of new dimensions in metric spaces of the sub-spaces of totality, with increasing degrees of small-world-ness (and even scale-free ness) of the networks. This in turn triggers the creation of new information (nodes) i.e. establishing connections of completely isolated nodes with the rest of the connectionist totality. Further establishment of connections leads to the evolution of the nodes i.e. information content and capacity (of the nodes) themselves. "Information seeds information".

This connectionist, irreducible, inherently random, open and recursive, emergent nature of totality leads to another conjecture regarding "the computability of totality". Finiteness of anything, say any sub-space of totality imposes boundary conditions or bounds on the subspace inducing a periodicity within the sub-space to accommodate the (perpetual) evolution (i.e. "change") taking place within that sub-space. This periodicity is reflected in the patterns of the sub-space. The patter represents a logical relativism. The pattern represents entities and their logical relationships. This pattern of a bounded sub-space is algorithmic and thus compatible. But, when the sub-space is not bounded, and is inherently random or strongly irreducible, then the pattern is non-computable. Thus, totality is non-computable. But any bounded or finite sub-space of it is computable or algorithmic.

Another conjecture I wish to forward in this paper is that of “Mesoscopic connectionism”. Within the complete connectionist representation of totality, this conjecture appears to be have the potential be significantly consistent. The connectionist nature of totality imposes strong, medium or weak interdependencies, - based on the degree of indirectness of the connectivities between entities. Also, with the perpetual recursive evolution of information, distributions of information changes over the sub-paces of totality, either linearly or non-linearly. New dimensions emerge with influence on existing dimensions. This, I feel, has the potential to create distributions, which are fundamentally mesoscopic in nature. This is an open and weak conjecture needing critical analysis. But human history has evidence of existence similar congruent ideas e.g the “The Golden Mean path of Life”.

Within the complete connectionist representation of totality and its sub-spaces, fundamental aspects that characterize the complexity of sub-spaces are the global holistic distributions within across subspaces, ensembles of sub-spaces, observer and observed sub-spaces, dimensionality of the metric spaces of sub-spaces, dependencies and connectionist structure within and across sub-spaces, their reducibility, informational capacity and corresponding fluctuations, linear or non-linear structure of phase spaces of distributions, and nature of the evolution of the sub-spaces. These distributions characterize the fundamental complexity of sub-spaces of totality to a significant extent.

Life or living world or biological world with biological systems, processes and phenomenon is probably the best example mirroring and inherently representing the “type” of all the above concepts of complexity because of the nature of life and living world with -species, taxons, kingdoms; information system structure of genes and genomes; the processes of molecular (e.g. evolution of genes and genomes) and organismal evolution; gene families, hierarchy and resolution of functionals; composition of the environment or biosphere and ecology; the connections between organisms, functionals, and biosphere; perpetuality of evolution of the living world and the biosphere and ecology (i.e environment) and the corresponding connectionist nature within and across themselves; embedded or “inherent” randomness in systems, processes and phenomenon of the living world- both inter and intra cellular and organsimal; uniqueness yet commonality, unity and diversity; difficulty to create and survive and propagate completely artificial life; evolution and existence of extraordinary or extreme forms of life; and phenomenon of consciousness.

Integration and Elaboration

John Archibald Wheeler once asked “The question is what is the question?” By this, one thing he meant was that there is serious doubt whether our approaches to understand reality and ultimate truth or nature of Totality are valid or not.

The objective of this chapter is to integrate all the ideas and work presented in this paper in a single framework, esp. to present a coherent description of the underlying intricate concepts of the work, and also elucidate the totality of ideas within which this paper should be perceived. The work embodied in this paper transcends but includes ideas, concepts, methods and tools from engineering and sciences and philosophy- it’s more of a kind of an attempt towards an “integrative philosophy”, reflecting convergence of thought, esp across disciplines, traversing from raw data to philosophy of Totality. It reflects a segment of a journey into the unknown and undefined, containing seed of ideas and concepts on complexity and understanding of totality, including the comprehension of universe, life and consciousness. Having said this, it must also be mentioned that new ideas are always difficult to communicate to an expert but general audience. New ideas suffer from a general drawback of not being unambiguously clear in the mind of the creator(s) (suspicions arise when it is claimed otherwise- neurobiologist would support this!).

The incomprehensibility of our world i.e. nature and totality (universe, life, consciousness and other entities) is a major concern for the pursuit of knowledge and understanding. There are several roads to reality. Among them, the concept of complexity is one such, although the concept itself is not defined, but people who view this world from the perspectives of complexity agree on a common underlying sense of the concept of complexity. What that is, is difficult to define, but a significant aspect of it being difficult to define is amply evident from the work embodied in this paper. Here, in this paper, I propose some ideas and concepts on Complexity. Having said this, it must be mentioned here itself, that, our philosophy of modeling, which have historically been kind of reductionist or rather simplistic, will not scale up to generate models of totality any significant sub-space of totality. This will become clearer from discussions that follow. Also, the study of reality- of real systems, processes or phenomenon and the philosophy of complexity have an extraordinarily wide gap in between, primarily because of the non-scalability of historical modeling approaches to the modeling of totality or rather any complex sub-space of it. In this context, one major objective of this paper is to attempt a projection of extant reality (and/or its complexity) to the level of totality or any complex sub-space of it, through the philosophy of complexity being proposed in this paper. The aim of this paper is to approach the incomprehensibility of totality from some new approaches and subsequently provide cases to illustrate the essence of those concepts. The concepts are simple yet intricate, so much so, that the cases provided to illustrate them may or may not appeal to any particular individual, but will have a basic validaity with anyone who sees through them. The paper illustrates a theory, a way to apply the theory, and the theory as a method for understanding extant problems and the type of inferences available from this method. The work of this paper also proposes directions towards a linguistic framework to which problems and inferences may be projected or mapped to create an ontological framework of advancing the theory of complexity. The work raises issues of engineering methods and protocols (here, computational in nature) to comprehend complexity, and how to design them towards solutions of utility.

I start with my conjecture, that the only representation of Totality is the connectionist representation i.e. nodes and edges or entities and relationships as complex network. This conjecture can be broken into two parts: one, that, there is atleast one connectionist representation of any part of totality and two, that all representations of totality are connectionist in nature. The first part of the conjecture implies the second part because of the word “any”. Thus, both these together say that the one and only representation of totality is the connectionist representation (Connectionism). Lets call this the “Complete Connectionist Conjecture” to identify this concept by a single phrase in subsequent discussions. Let us mention some other aspects relevant to the conjecture. The act of measurement or observation creates duality- of logical positivisms and consequently logical negativism, based on whether you can measure or observe something (observability) or what is the result of the measurement or observation. Controllability imposes another level of duality, according to the same reasoning. Change, which is known as the “only thing constant in the totality”, has relation to duality. Nonlinearity, in deep sense, is a creator of dimensionality -the creator of new dimensions in totality. The complete connectionist conjecture, stated above can be understood as a form of relativism, where all entities are related to other entities through relationships. This concept of relativisms can resolve the dilemma of holism and reductionism. In reductionisms, when anything is considered to be composed of something else or anything can be explained in terms of something else, then it creates intractability in defining holism. But both reductionisms and holisms can be accommodated within the the concept of relativism in the complete connectionist representation of totality.

Information evolves. It evolves by making new interconnections with existing information. This is accommodated within the complete connectionist representation of totality where an entity or a new node gains existence i.e. meaning, when it establishes connections with atleast one of the existing nodes or entities. If this new node is a quantum of information (or knowledge), then it gets defined only when it establishes connection with existing nodes. The connectionist nature of totality triggers, through the interconnections of totality, gradient in information, which triggers the evolution of information and its subsequent creation, or space for its creation. This leads to the infinite recursive Information, where information becomes a completely open system, perpetually evolving. Thus the entire connectionist representation of totality, a complex network, evolves. Transformation of information- with evolution of nodes i.e quantum of information establishes new interconnections- thus achieving new meaning- thus getting transformed. Thus evolution, both, creates and transforms, information. Evolution transforms information either partially or completely (some interconnections changes OR all interconnections changes- with other nodes). This can also happen- creation or transformation or both- to entire sub-spaces. Identical or near identical information (in terms of connections i.e. similar or identical connections) can arise by completely different processes of evolution of information e.g. identical information evolving in different subspaces of the connectionist totality due to highly non-linear processes of evolution – thus leading very interestingly to dimensionality reduction. So evolution of information also triggers dimensionality reduction. Apparently, the probability of this event seems to be very less considering the enormous dimensionality of the metric space of the connectionist totality or even a large subspace of totality. Processes of evolution of information in the connectionist paradigm are similar (even, at times nearly-identical) in neighborhood of metric spaces leading to evolution of identical or near identical information in a micro-subspace of the neighborhood metric space of node (of information). Critical transitions effecting establishment of connections between distant (in a connectionist way) subspaces leads to the creation of new dimensionalities or changing the distributions of the states of a subspace. Local processes of information evolution, esp. linear ones, operating in neighborhood of nodes or subspaces creates similar informational subspaces, primarily changing the distribution of the states of the containing subspaces. It is evident that in this connectionist structure of totality, single events of creation of information (nodes) potentially significantly increases the complexity of subspaces, esp. due to the creation of new domains in the phase space of the combinatorial complexity of interconnections or relations. More linear evolutions esp. from small neighborhood metric spaces maintains symmetry between nodes (quantum of information) which thus become evolutionary connected. Survival of nodes (i.e. interconnections of nodes with the rest of the connectionist totality) depends upon the information content or identity of the node which determines whether it can form interconnections with other nodes and becomes stable (a form of computational complexity of each node). The same goes for entire sub-spaces. If the information content is not adequate for establishing connections with other suitable nodes, then the perpetual evolution of information over the complete connectionist representation of totality delinks the node from the network, thus isolating the node and thus nullifying its information.

This, overall, bring us to the question of generalized laws of totality- some call it ultimate truth or reality. The complete connectionist representation represents an irreducible complexity of totality, because due to the connectionist nature- totality cannot be reduced to a single node or quantum of information and thus an inherent randomness within totality (since something which cannot be compressed beyond a finite point must have some inherent randomness embedded within it.) This irreducible complexity and inherent randomness of totality makes the computability of totality infeasible. But since finiteness makes spaces computable, because finiteness or boundary conditions induces periodicity and thus pattern (logical relativism), it makes the space computable. Thus this entails that totality is by itself non-computable, but any sub-space thereof, which is finite, may be computable. This also entails, that the subspace may be represented as deterministic without inherent randomness when observed from a space or reference outside that subspace. The computable subscape embeds an inference tree in connectionist representation of totality. This raises intriguing aspects of finites and infiniteness. As the totality of information space evolves according to the complete connectionist representation, it may add new dimensions to metric space of connectionist representation of totality. Nonlinearity, in a sense, has role in triggering criticality or singularities and in the same way creating new dimensions in the metric space of totality. This happens when connections are established between nodes that are not neighbors- more distant nodes when connected has the probability of triggering the emergence of new dimensions of the metric space of the connectionist totality. Linear evolution of information occurs when connections are established between neighborhood nodes (nodes in the neighborhood of metric space- information evolves by a “sequence” of steps connecting neighborhood nodes, nonlinear evolution of information occurs when distant nodes- distant in terms of connectivity get connected). It is possible that new dimensions evolve in such ways, and if it does so- i.e. new dimensions are added, then probably generalized laws of totality are intractable to achieve, even if they incorporate randomness to accommodate the inherent randomness of totality. Thus, totality moves towards irreducibility. When/where new dimension will emerge is extremely difficult to predict because of the complete connectionist nature of totality- where the entire network is colossally complex. The number of dimensions needed to describe a system or a process is an indication of its complexity.

The act of measurement or observation creates duality- logical positivism and logical negativism, and thus creates gradient of interactions amongst the entities in the connectionist totality and creates further information. This is equivalent to an uncertainty principle of totality within the complete connectionist representation of totality by which, the relativity (the TOTAL amount of information) space of information changes during the act of measurement or observation. This in effect changes the meaning of information obtained by the act of measurement or observation.

To compute totality or any subspace of it, a finite universal semantics must be the representational language of the complete connectionist totality. This would be an universal ontology of totality. This ontology, I would suggest be of a generalized complexity theoretic ontology of totality. The major purpose of the universal controlled language or semantics for any system, process, phenomenon- either physical or mental is to universally converge to a finite set of terms/phrases that would define the set of all properties possible for any systems, process, phenomena- either physical or mental. This would allow us to think of any sub-space of totality in terms of a universal language. This would allow for the comparison of all forms of trends and events of any systems, process, phenomena- either physical or mental, in totality, at all scales of abstractions, on an identical platform. This is not possible as of now because no such language exists today. Besides ensuring the computability of any subspace of totality, this would lead to comprehensibility of totality. This ontology of complexity would also help to determine the points of irreducibility of totality and the seat(s) of randomness in totality. This universal ontology is essential to decipher common laws across different sub-spaces (e.g. systems, processes or phenomena at various scales of abstraction and projections), which is one of the basic aims of complexity. This universal ontology enabling the computability of (major) sub-spaces of totality leads to the definition of informational nodes through establishment of connections (the process of computation) with the rest of the totality. This would lead to a computable representation of the phase space of the complexity of totality or any of its subs-spaces.

In the complete connectionist representation of totality, to compute through inference trees, certain finiteness must be imposed to extract out the subspaces of interest. These subspaces are understood in terms of their relativism with other subspaces. Since, measurement from outside a subspace corresponds to an observer, therefore each of these subspaces are reference frames, sometime overlapping i.e. with dependencies (probably inherent) and sometimes not (weak dependencies or mutually exclusive metric sub-spaces). The inherent dependence or relativism i.e. connectionism between these reference frames or subspaces, gives the potential for evolution of information, with or without new dimensions, through the interaction gradients of these subspaces/reference frames. The dependencies also express one subspace in terms of another and thus determine regimes and degrees of irreducibility, if any, within the containing metric spaces of those subspaces or reference frames. One or more subspace can be represented as the environment of any subspace (system).

The metric space of each of the subspace, and their dimensional transformations (new metric spaces with transformed dimensions or reference frame) gives the representation of one sub space with respect to another. The properties of subspaces of each of the subspaces, the relations or dependencies between the smaller subspaces with the subspaces in terms of the observer subspace provides the connectionist complexity of the observed subspace and thus the symmetry and its degree of information of the nodes of the observed sub-space in terms of the observer subspace. The dimensionality of the observer and observed subspace creates the complexity of the dependencies of the information of one sub-space in terms of the dimensionality of the other. This demands the examination of the dependencies in various sub sub-spaces (projections in metric spaces with combinations of dimensions of both observer and observed subspace) to understand the realms of the metric space they (dimensions of observer and observed sub-spaces) occupy. The representation of an observed space in terms of an observer space also elucidates the emergence of (new) dimensions, if any, with the sub-space of the union of the observed and observed space. As mentioned earlier, emergence of dimensions makes generalized laws practically impossible, as the evolution of information, once triggered, is a perpetual phenomenon.

Now, let us consider the enormous ensemble of sub-spaces of the connectionist totality. The state (or state vector) of the observed sub-space within the complete connectionist evolvable totality, and the states (or state vectors) of the sub-spaces of the observed subspaces in terms of the observed subspace (and its dimensions), their own connectionist structure and dependencies (e.g. correlations) within and across observed subspaces characterizes the connectionist complexity and evolvability of the observed subspace and its sub-sub-spaces. The dimensions of the observer sub-space, being present in the complete connectionist totality, can be resolved from with simple (probably linear and/or uni-dimensional) subspaces, representing the neighborhood metric space of those dimensions within the observe subspace. Thus, accordingly, the states (or state vectors) lie within a neighborhood (allowing fluctuations) of the states themselves. The distributions and dependencies, symmetries and complexities, the conservation and divergence of the states/dimensions themselves and the sub-spaces they compose represents the complexity, of the states/dimensions themselves and the subspaces they compose. Deciphering the states/state vectors, or more fundamentally their symmetry and complexity, their conservation and divergence gives an understanding of the complexity of evolutionary process itself and also fundamental design complexity of the spaces/entities generated or created. The nature of the distributions and dependencies of states indicating a global picture of the connectionist subspace, maximum reducibility of holistic dimensions and basis or transformed dimensions in metric sub-spaces, symmetry or sharing of dimensions among sub-spaces, (e.g. sharing of communication dimensions indicates how similarly can environment affect different sub-spaces/systems), the total dimensionality of any sub-space/system -indicating its combinatorial complexity and evolvability (phase space), how much bandwidth of the spectrum of connectionism it interacts with and thus influences the complementary totality, are some of the fundamental aspects of the complexity of sub-spaces of totality. The information capacity and the structure of the space of information capacity indicating any biasing towards any particular form of information capacity of the dimensions of the observer space signifying a form of informational identity of the dimension (or node in the connectionist representation), weightage of the dimension in the complete observed subspace, weightage of the dimensions in major sub-spaces of the observed space, the universality of the dimensions of observer space across the bounded/ finite micro-subspaces of the observed space indicating the span of the metric space of the observed space dependent upon/connected to this dimension, their weightage in those micro-subspaces indicating any biasing within the universality spectrum of the dimensions amounting to robustness, dynamical role or structural role, and spectrum resolution (due to evolvability of the individual dimensions), dimensional transformations and corresponding metric spaces of the observed space in terms of transforms of the dimensions of the observer space indicating the dimensionality in a different but may be more fundamental space, the state vector of the original dimensions in terms of the transformed dimensions indicating a more fundamental nature of the previous (“holistic dimensions”), spectrum resolution of the holistic / original dimension in terms of the new metric space (indicating potential for further connectivity and thus evolvability of the subs-space), the complexity of the spectrum – representing each dimension from the observer space as a sub-space by itself, are some other fundamental measures. Several of these aspects are exemplified by distributions, esp. of states or state vectors and their components, in terms of possible nearly orthogonal or ortho-normal basis spaces. From the above discussion, it is also clear that the complete connectionist representation of totality also weakens the concept of existence of model or representative sub-spaces, esp. when sub-spaces are represented in terms of defined states or state vectors (and their components or dimensions in the covering metric space). Having said this, it must be said that (degrees of) symmetry pervades (or rather must pervade) the complexity of connectionist totality. This is exemplified by the sharing of dimensions (of metric spaces) of observer sub-spaces across the ensemble of observed sub-spaces. This, in all fundamentally signifies the equivalence of symmetry and complexity. Within the connectionist conjecture, once a space e.g. totality, is irreducible, it implies inherent dependencies and/relationships between the entities in the maximally reduced sub-space. This dependencies and/or relationships are the origins of symmetry or invariant transformations or conservation across sub-spaces. Thus, the complete connectionist representation of totality induces a dependency and relativism across all sub-spaces and/or nodes- this effectively induces the irreducible complexity of totality. But on the other hand, this relativism of sub-spaces of totality is the origin of symmetry within and across sub-spaces and thus and underlying symmetry across the entire space of totality i.e. super-symmetry.

Several aspects of the connectionist (complex network) sub-space of totality represent the complexity of the sub-space. As stated earlier, the integrative inference of these aspects over the totality of the sub-space represents the complexity of the sub-space, both as a whole (holistic) and as individual aspects of complexity of the sub-spaces. The complete integrative inference is (only) possible when the objectivity of the individual aspects of complexity are integrable over the totality of the sub-space. Below we delineate several aspects of the complexity of the sub-spaces. The states and / or the dimensions of sub-spaces, their distributions and dependencies within and across observed sub-spaces characterize the complexity and evolvability of the subs-spaces. The dimensionality of the metric space of any sub-space is representative of the complexity and evolvability (realms of phase that are yet to be traversed by the evolving sub-space). The density or sparse-ness of sub-spaces, delineation of the identity e.g. functional i.e. dynamical identity of sub-spaces, density, sparseness and distributions of dimensions per observed sub-space or its constituting sub-spaces thereof are other aspects of the complexity and evolvability of the spaces. The range of the distributions indicating the boundaries of the states and their corresponding sub-spaces in terms of their dimensions, relative complexity and/evolvability of certain well-characterized sub-spaces in terms of the entire space of the observed sub-space, density and sparseness of dimensions in major or minor sub-spaces of the observed sub-space, esp. those identification of sub- sub-spaces defining the range of the dimensionality or its corresponding density, the complete dimensionality of the observer space for the fundamental transformations of the metric space of the observer space, indicating the potential evolvability, dimensionality and connectivity of individual nodes or entities or sub-spaces of the observed sub-space, are some other fundamental aspects of the complexity and evolvability of the connectionist sub-spaces of totality. Besides, ranges and distributions of the strength of the dimensions of the observer space within the complete observed space, indicating the probability or potential fluctuations in the actual identity of individual dimensions of the observer space, symmetry or degeneracy in the dimensional transformations between the dimensions of the observer metric space, ranges and distributions of maximal information content of the nodes or entities of the observer sub-space being representative of the structure and evolutionary dynamics of the informational space of each node or entity of the observer sub-space, degrees of linearity and non-linearity, dense and sparse regions of the phase space of evolutionary dynamics of the maximal information content, symmetry or degeneracy in the ranges of maximal information content of each node or entity of the observer sub-space are some other aspects of the fundamental complexity and evolvability of complex sub-spaces. Certain other significant features of complexity and evolvability of the complex sub-spaces of connectionist totality are the absolute weightage of the dimensions of the metric space of the observer sub-space within the observed sub-space, density or sparsity of nodes/entities of the observer sub-space in the ensemble of observed sub-spaces, dimensionality of each of the subspaces in the ensemble of the sub-spaces and transformation of the ensemble of observed sub-spaces in terms of dimensionality of the each sub-space in the ensemble, into connected sub-spaces in terms of the dimensionalities of each of them, ranges and relative dimensionality of observer sub-space constituting coherent sets of observed sub-spaces and relative dimensionality of these observed sub-spaces. Some other important aspects representative of the complexity are the compositional density and its distributions of the observed sub-spaces in terms of the nodes/entities of the observer sub-space, dependencies between the compositional density and dimensionality of the observed sub-space in terms of the observer sub-space, combinations of observed sub-spaces and the unified symmetry of dimensionality of the sub-spaces per combinations as a whole, its degree, symmetry of combinations of observed sub-spaces among themselves with respect to the dimensionality of the observer sub-space, absolute distribution and range of the symmetry of dimensionality in total space of the combinations of observed sub-space, total dimensionality of sub-set of observed sub-spaces, its ranges and distributions, patterns and degrees of connectivity, with respect to symmetry of dimensionality, of observed sub-spaces or its subsets with other observed sub-spaces in the ensemble of sub-spaces. More aspects representative of the complexity of sub-spaces are generality or universality of dimensions of the metric space of the observer space within the ensemble of observed-subspace, transformations of subsets of the ensemble of combinations of observed sub-spaces in terms of the degree of symmetry of dimensionality, distribution of the degree of symmetry of dimensionality of observed sub-spaces with complementary members of the ensemble of the observed sub-spaces, and sub-space specific dimensionality the ensemble of observed sub-spaces. Representation i.e. states or state vectors of ensemble of observed sub-spaces and their combinations and their original dimensions in the metric space of dimensions of transformed (may be more fundamental) observer space and the distributions of the transformed dimensions in observed sub-spaces, exclusive dimensionality of observed sub-spaces and their combinations, dimensionality of original dimensions of the observer space in terms of the dimensions of the transformed observer metric space, symmetry of dimensionality of original dimensions of observer metric space, distributions of dimensions of transformed metric space in the ensemble of original dimensions of observed sub-space, degree of symmetry between dimensions of one observer space with respect to dimensions of a transformed observer space, and of course symmetry between exclusive dimensions of observed sub-spaces or their combinations of observed sub-spaces are some of the complementary aspects of complexity and evolvability, which together with the all the above aspects form, to an significant extent, the almost complete “complexity map” or rather a fundamental architecture of the complexity and evolvability of the complete connectionist representation of sub-spaces of totality. Ofcourse these aspects of complexity may not be all independent of each other, but they approach a kind of a fundamental set by their definition, but also proof of their independence is substantially difficult, from very nature of the concept of complexity being propounded in this paper.

Ab-intio with no a priori hypothesis discovery science based complexity theoretic studies leads to the actual characterization and the nature of the variations within and across connectionist spaces or networks, thus enabling asking of meaningful questions not possible through conventional a priori-hypothesis based studies. Such efforts could be characterized as providing the building blocks or raw materials out of which hypotheses can be formulated—metaphorically, words of a universal “language” for expressing hypotheses. The availability of knowledge of large sub-spaces is expected to enable universal questions to be addressed globally following the approach of discovery science and/or exploratory analysis. When subspaces in connectionist totality evolves with an inherent randomness embedded in the process of evolution (dependent upon their position in the connectionist totality), then the final topological organization (and their corresponding dynamical architecture) of such subs-spaces are unknown and undefined. This is esp. true when the degree of irreducibility (also dependent upon observability and its degree) and thus embedded randomness strongly influenced the evolutionary dynamics. Which solution sub-spaces has evolved, from a large phase –space of solution sub-spaces is almost impossible to ascertain from abinitio apriori hypothesis based approach. Thus, the variations and correlations of the observable states/state-vectors and their dependencies or relations within and across themselves will elucidate the dimensions, dependencies and laws (their number and nature e.g. dimensionality) governing the sub-spaces. The ab initio discovery science method is suitable because it does not fundamentally assume any a priori connections or relations between nodes or sub-spaces (hypothesis or law) and thus eliminates the choice of incorrect observer sub-spaces or reference frames to characterize the complexity and governing laws of complex observed sub-spaces, esp. those evolved through evolutionary dynamics with embedded randomness. Deductive a priori hypothesis driven analysis seems to be playing with trial and error to understand such colossally complex adaptive evolved sub-spaces as they may (if simple) or may not (if complex) enable their definitions or characterizations/understanding. But whether they are simple or complex or colossally complex is unknown before the characterization. But ab-intio-no-aprori-hypothesis-based complexity-theoretic approaches are general and can be applied to any sub-space- simple or complex, deterministic or stochastic or any combination of these thereof. But the main limitation of this approach is the limited degree of observability of states, and degree of computability of the connectionist sub-space (due to the colossal complexity), which limits a complete complexity theoretic analysis of connectionist sub-spaces of totality. But classifications of sub-spaces (systems, process and phenomenon) based on the pervading symmetry (discussed before) across ensembles of sub-spaces enables the transformation of the connectionist totality to a “reduced connectionist totality” thus enabling a hierarchical representation of the connectionist totality with reference to chosen observer spaces. The levels of classification hierarchy enable the justified selection of some suitable sub-spaces or reference frames which have necessary (sometime sufficient) degrees of observability and computability providing substrates for illustration of such ab-initio- no-apriori-hypothesis-based discovery-science complexity theoretic analysis.

In the connectionist representation of complex subspaces, represented by innumerable nodes & edges or entities and their relationships/interactions, the causal structure (cause-effect relationships between the entities) of the sub-space becomes obscure and thus irrelevant. This does not imply that there may not exist any pure-ly causal structure. This may exist, esp. depending upon the (observer) sub-spaces. The distributions of states/state-vectors indicate the linearity or non-linearity within the state implying whether the state has undergone linear or non-linear evolutionary dynamics. Minimal variations indicate that either weak dynamical processes of evolution of information has affected such states or entities/nodes/sub-spaces, or otherwise strongly but nullifying dynamical process of information evolution has influenced that state/node/sub-space. Weak correlations/dependencies also generate the possibility of multiple laws affecting the interactions between states/nodes/sub-spaces, while strong dependencies raise the possibility of single or few laws (connections) governing the interaction between the states/nodes/sub-spaces. This characterization also elucidates the symmetry within the complexity of the spaces. The uncertainties of totality are part of the complexity of totality, and the degree of uncertainties (or irreducibility) and their contexts (actual and neighboring connections or sub-spaces) are necessary information for deciphering the nature of the laws (global connections) affecting the complex network. The symmetry of laws (global connections), metric-space dimensions and their dependencies is the super-symmetry in the totality (within the context of complexity). Various laws (e.g. global connections) are triggered at different levels of abstractions (from various sub-spaces), and this can be a very fine point of triggering which cannot always be observed or computed. If a matrix or vector be constructed of all the observables of a complex network, then the distribution and degree of correlations will elucidate the number of laws governing such systems and the degree of irreducibility of such systems. Thus must be interpreted in the context that variations and correlations in a connected system gives the information content, degree of compression (into laws) and the degree of randomness and the indication of the independent set of laws or dimensions that have made the system. The lack of dependencies between the parameters is as important as the dependencies in the theory of complexity (as being propounded here) as this is important information in the holistic context of understanding the complexity- “what is present becomes information only in the knowledge of what is absent”. Since information is relative to the totality or any sub-space thereof, so what connections are present and what are absent both has to be taken into consideration to establish the informational identify of a node or entity or sub-space. Recent research into complex networks, generalized models of complex systems in terms nodes and edges, have indicated my long standing conjecture- about the nature of the world being fundamentally determined by the nature of connectivity (and not the exact connectivity) – the nature of the connectivity being the prime determinant of the dynamics of complex networks or systems (rather than the nature of the elements in terms of their basic linear or nonlinear behavior) being probably true. [Arenas et. al. Physics Reports, 469, 93-1153 (2008).]. But, having said this, it must also be categorically mentioned that it will be very difficult to even appreciate the value of the results obtained by complexity theoretic approaches if we continue to think in non-complexity theoretic process. Complexity theoretic approaches also help to uncover underlying unrecognized principles. But it must be appreciated that discrete and defined conclusions, as prescribed the historically dominant philosophy of science, should not be searched for. The control information, as by the historically dominant philosophy of science, will be available only after several exercises of similar kind. This will create a platform of conventional scientific inference from a population of possible incomplete inferences. This approach is difficult to avoid, as per our current understanding of complex systems. But one major conceptual issue that needs to be resolved is the representation of the complexity of the sub-spaces or complex network e.g. the distributions of the states or state vectors. This issue primarily arises due to the extremely diverse nature of the components of totality e.g. continuum, discrete and enumerative nature of the components and thus the complexity phase space of totality or its sub-spaces. This makes it difficult to give unified representation of the distributions of the states of nodes/subs-paces unless ofcourse the nodes or entities of the sub-spaces are of unified nature. Thus one easy way out is a qualitative, fuzzy, context-dependent representations which is somewhat universal to an extent of consistency e.g. representing the distributions in suitable ranges of their values. In all, the discussions above, demonstrate the validity and justification of ab-initio no apriori hypothesis , discovery science based complexity theoretic approaches to study sub-spaces of totality.

The nodes or entities or subspaces comprising any large sub-space of totality, e.g. in the context of biocomplexty or the complexity of life/living world is composed of widely different “types” of nodes/entities or sub-spaces. Also, each node or entity or sub-pace location within the large sub-space of totality or totality is itself defines the objective identity of that node or subspace. Due to the complete connectionist and evolving nature of totality, closed, complete and objective identity or measure of the node or sub-space is computationally intractable. Thus, within this concept of complexity (as being proposed in this paper), it is meaningless (since computationally intractable), to define absolute measures of complexity of the nodes or sub-spaces. Since this is true for any sub-space or node, thus, defining closed, complete and objective measures of any major sub-space of totality is also computationally intractable, either in terms of the complementary sub-space of the major sub-space or in terms of the sub-space itself, because the objective measures of the sub-set of the sub-spaces are themselves computationally intractable, thus a total measure is also either completely non-integrable or may be to an extent only partially integrable. Thus to characterize the complexity and corresponding evolvability of a sub-space, enumerative esp. qualitative or semi-quantitative measures are the best representation. The enumeration or distribution, being that of an open evolving phenomenon, is not objectively bounded with respect of the totality. Thus delineation of the actual distribution or enumeration is one way of overcoming this issue. Since the method of comprehension of totality or any major sub-space thereof or even a node in terms of the complementary totality is limited by the weak understanding or measures of nodes/sub-spaces, uniqueness of the nodes or sub-spaces, non-computability and non-integrability of the measures, the solution is to comprehend any major sub-space of totality (in totality i.e. holistic way) as the best estimate, however “incomplete” the best estimate may be, but comprehending the best estimate a partial inference of the sub-space of totality. Thus, the diverse (e.g. continuum, discrete and enumerative) nature of the complexity phase space makes it extremely difficult to give unified representation of the distributions of the states of nodes/subs-paces. This leads to the need of some qualitative, fuzzy, context-dependent but somewhat universal representations incorporating ranges and bounds. In this context, one pragmatic but strongly valid way of characterizing the complexity esp. in the holistic sense of a complex space is to elucidate the nature of the distributions indicating a global picture of the connectionist (sub) space. The distributions, by their definition, take into account the characteristics of complete space to define itself (the distribution). Thus qualitative context-sensitive fuzzy distributions for deduction of best estimates of the aspects of complexity of spaces is one major way of approaching these problems.

Thus, from all of the above, integration of diverse plausible inferences on complexity of sub-spaces is required to achieve a more holistic understanding of the complexity of the sub-spaces.

Complexity i.e. the picture of the incomprehensibility, complete connectionist representation of totality, embedded randomness (esp. during evolution) and the infinite recursive evolution of information demands a suitable set-up that mirrors the theory. Life, or the living world is the best set-up that mirrors the concepts stated above, although traditionally due to the nature of pursuit of biological knowledge, the living world is maximally far off from the concept of complexity as embodied here. The parallels between the complete connectionist representation of totality and that of life/living world are manifested by the apparent irreducible complexity of life, with some inherent information, noise and or randomness underlying the laws of life. This is also manifested because there is a critical distinctiveness between the eukaryotic sub-spaces and prokaryotic sub-spaces- major sub-spaces or kingdoms of life, where within prokaryotic systems archaeal sub-space are exotic--- so this is a very interesting subspace of totality having unique and exceptional structure and thus, probably dynamics, so to cast our paradigm to this subspace was natural to demonstrate the concepts being proposed here. The inferences from the integrative meta-philosophy of complexity, esp. from living kingdom and archaeal kingdom will provide information about complexity of components responsible for connectionism, how levels of abstractions can be traversed and their complexity revealed which provide information about the nature and origin of connectionisms and the evolution of such complexity e.g. computation capability or connectionist capability. How such integrative meta-philosophy and approach of complexity can uncover inferences about a complex subspace of totality, while also being consistent with fundamental extant knowledge but also finding regimes of incomplete information avenues for further questions is something that needs exemplification. Thus, it is itself a task to represent the problem from the living world in terms of the theory of complexity as embodied here. But this was one of major the objectives of this paper (another one being to make it feasible to study or investigation after successful representation or casting of the problem onto complexity theory). Consider the living kingdom, at any level, a hugely complex network made up from innumerable nodes and interconnections and an equally diverse types among the nodes (entities, parameters, elements etc) and their interconnections (relationships, interactions etc) (where sometimes the interconnections are nodes and the nodes are interconnections). This would represent the MAP of the living kingdom in desired sub-spaces or reference frames. Here the systems-subspaces (e.g. species) and environment-spaces (biosphere, ecology) are intricately coupled. Living things interacts with its environment through communication manifested by the signaling dynamics of the components of cell i.e. proteins. Cells react to its environment according to the signaling and communication proteins or components present in it. The cell therefore has to integrate the information coming from separate signals so as to make an appropriate response. Signaling, between systems and their environment is the prime factor in the adaptive process of evolution of these living solutions to the dynamically changing environment, since signaling is the mechanisms by which the solutions interact with their environment or objective function, within the solutions themselves as other solutions are also composing the environment or total objective function (the environment being composed of a part by itself and of solutions to the objective function, where each of the solutions influence both the inherent component of the environment also the other solutions which form part of the total environment or total objective function for a particular solution. It is also evident that the particular solution which is adapting to the environment is also influencing the environment.), and also each solution maintains itself with the process of internal signaling. Signaling and communication networks are also responsible for generating the unity within diversity of living kingdom. The signaling and communication systems establishes the necessary message passing and signaling between the controlling parts and also between the engine and the controlling parts for accurate and efficient transfer of signals from one part to another, and its interactions with its environment. Thus, connectionism is established in the living world through the signaling and communication esp. by the role of (protein) components. Thus, the infinite recursion of evolution of information nodes/sub-spaces e.g. species, communication systems, signaling components, is triggered. The role the entire biosphere plays in this complex evolutionary process is observable through the constraints set by the bounded yet changing environment and the network dynamics of the organisms within and across themselves to create the ecosystem, which, in turn, affects the process of evolution- a cyclic and recursive phenomenon esp. manifested by signaling and communication between the species sub-spaces and the environment through signaling sub-spaces. This leads to the emergence of the inherent property of adaptability within the complex uncertainty of the environmental constraints or demands. The nature of the evolutionary process seeded with a population of entities and exchange and/or addition of information (and probably noise too) over local and global scales both within and across individuals is the prime reasons responsible for the unity within diversity in the living world. Within the connectionist paradigm, this leads to the evolution of the complexity of the sub-spaces, which are interconnected amongst each other. The colossal complexity of the sub-space of the living world/life thus becomes highly irreducible within the covering space and thus embeds randomness within the spectrum of connections. This leads to the generation of individualism where no two sub-spaces (cell, proteins etc.) are identical justifying that “every living thing is unique. This overall, raises epistemological questions about the nature of biology itself and on the question of how biological phenomena can be understood. When this perspective is considered for the biological functional dimension of signaling, which is the dimension responsible for the interactions of life, evolution an environment or in other word responsible for the interaction and thus stability of solutions with other solutions and the components of the total objective functions, and thus influences the process of evolution of the solutions themselves (as a feed back mechanism). This perspective becomes even more remarkable when such complexity is elucidated independently for the solutions or taxons which form a uniquely different class of solutions, extremely different when compared to the total set of solutions, i.e. archaeal kingdom and viewed in terms of the solutions themselves and wrt to the mainstream class of solutions. Archaea are unexpected form of life i.e. distant subspaces- very distant from the major sub-spaces or rather sub-spaces on the “boundaries” of the connectionist totality space. Archaea are very different from our usual concept of life esp. with them inhabiting extreme environments thus having completely different lifestyles than conventional known life forms, thus giving a new but also critical dimension to life or living world. There is a major eukaryotic and bacterial component in archaea, yet they are bordering the sub-space of life, having some connections or relations with far-off sub-spaces, with some dimension common with some subspace and some other dimensions of different informational identity symmetrical to other sub-spaces. This makes the archaeal kingdom is uniquely different from the other two kingdoms. This extraordinary and intriguing uniqueness of the archaea immediately compels us to focus on its signaling system, to see what kind it is, how much it is similar to that of the other two kingdoms of bacteria and eukaryotes and how much it is different from them, how much the dependencies are similar and different within and across the kingdom or lineage specific- to establish the complexity and evolvability of the archaeal sub-space from the observer sub-space of signaling and communication i.e. the sub-space responsible for connecting the archaeal sub-space to the complementary sub-space of life and living world and characterizing its complexity and evolvability.

When analyzing complexity and corresponding evolvability, one has consider two sub-spaces of complete connectionist totality- one the observe space and the other the observed space. These two sub-spaces are “complementary” to each other. But due to the irreducible and open nature of connectionist totality, any sub-space can be either of the two forms- observed or observer. In this context, to elucidate the basic nature of the philosophy of complexity (as presented in this paper), one must show the philosophy, theory and approach of complexity for both possible combinations of sub-spaces (observed and observer) to illustrate the theory of this paper. Thus in the context of the totality of the space of biocomplexity, two such sub-spaces can be identified at a computational tractable level of hierarchy – that of functional sub-space and organism-sub-space.

But to illustrate the nature of pervasive connectionism, such distinctions have to be made even among smaller sub-spaces of these two major sub-spaces i.e. of functional and organismal. This puts limits on computational tractability of the study of the problem. This leads us to the choice of only one functional aspect of one sub-space but the other organismal sub-space was taken maximally due to the limited number of organisms available at the time of the study. Signaling functional sub-space was specifically chosen because of reasons explained later. Thus, work in one chapter focuses on the detailing of the organismal sub-spaces in terms of two major and broad aspects of functional sub-space- holistic dimensions and unitary dimensions. Here signaling is the observer space and the organismal space is the observed space. While in another chapter, the work focuses on the detailing of the functional sub-space in terms of major and broad classifications of the organismal sub-space. Here, the organismal space is the observer space and the functional space is the observed space. The amount of data is just adequate at the time of the analysis to elucidate these aspects of complexity. (Subsequent work with newly added data has shown the significant increase of computational power and optimization of computation required to tackle such data). Increasing the amount of data will be representative of mechanistically larger but conceptually identical spaces of biocomplexity. These perspectives, in all, would give us the generalized complexity of the biological solutions and the functional dimensions and also of the major exceptions of the living world. The nature of biological complexity, in spite of being apparently strongly irreducible, will be elucidated gradually by this line of thought. I feel deep insights into new interpretations of life will be gained through these form of exercise as embodied in this paper and will hopefully be pursued by others who sees through the significance of this exercise.

Molecular Biocomplexity signifies the analysis of complexity of biological sub-spaces from the reference frame or observer sub-space of molecules. This level of abstraction is suitable for exemplification of the philosophy and approach of complexity (as propounded in this paper) primarily because data across different levels of abstractions and projections is very well semanticized, atleast when compared to other fields. This makes the complexity characterization computable or computationally feasible and thus makes the inferences of some consistent inferences plausible. Yet, the complexity of sub-space is complex enough which makes it difficult to perform such a study even at these among the highest levels of abstractions, esp. for life or living world whose complexity is colossal (dense connectionist subspace and densely connected neighborhood spaces). Thus it’s extremely difficult to choose a suitable path of analysis but yet this must be done. Research projects should attempt to put forward a philosophy of ab initio characterization of complexity of colossally complex systems like life forms and demonstrates a analysis procedures to tackle such complexity. This paper is an attempt to identify suitable tractable (esp. computationally) yet demonstrative sub-spaces. Thus here selective the levels of abstraction/sub-spaces of 3 kingdoms of life and the fundamental functionality of living systems e.g. control, communication, metabolism enables a semi-complete but an acceptable level of complexity theoretic analysis. Such analysis demand a level of abstraction that is high enough to form a complete set i.e. the 3 domains of life or fundamental functional subsystems/sub processes of living organisms. In the living kingdom, from a connectionist paradigm, there is a critical distinctiveness between the eukaryotic system and prokaryotic systems, where within prokaryotic systems archaea are exotic--- so this is a very interesting subspace of totality having unique and exceptional structure and thus, probably dynamics, so to cats our paradigm to this subspace was natural to demonstrate the concepts being proposed here. How such philosophy can uncover inferences about a complex subspace of totality but also being consistent with fundamental extant knowledge while also finding regimes of incomplete information avenues for further questions is a philosophical realm that has not been popular to pursue. When this perspective is considered for the biological functional dimension of signaling, which is the dimension responsible for the interactions of life, evolution an environment or in other word responsible for the interaction and thus stability of solutions with other solutions and the components of the total objective functions, and thus influences the process of evolution of the solutions themselves (as a feed back mechanism). This perspective becomes even more remarkable when such complexity is elucidated independently for the solutions or taxons which form a uniquely different class of solutions, extremely different when compared to the total set of solutions, i.e. archaeal kingdom and viewed in terms of the solutions themselves and wrt to the mainstream class of solutions. These perspectives would give us the generalized complexity of the biological solutions and the functional dimensions and also of the major exceptions of the living world.

Within the context of Molecular BioComplexity, the genome of an organism can be considered as an information system encoding the functionality of the organism, either completely or partially. The genome can be considered as the repository of the information for the molecular design and dynamics of an organism but how much of the information does the genome store is still a matter of debate. One among other existing views is that the information about the environment (that to which the organism has already adapted) and its dynamics is encoded in a transformed form in the genome and is later on transformed again into the molecular design and dynamics during the complete cell cycle. The processes of genome evolution which acts on these genomes are wide, varied and fascinating. The fundamental processes of molecular evolution influence the evolution of the genome and vice versa i.e. the evolution of the genome as a whole often influences the evolution of single genes or gene families since the entire genome acts a whole network, where one component influences the functioning of another. The genes at one level of abstraction can be considered the units of functionality. The process of genome evolution and the associated generation of new and expanded set of genetic material or genes/proteins imposes a connectionist structure on the protein universe in terms of homology (reflected as similarity at sequence level), where a protein member or node is connected to other node(s) directly or indirectly, forming a connectionist sub-space of the connectionist totality. The gene/protein families can be considered as the basis set of functions present in an organism whose components [like vector components] interact with each other to produce a functional dynamic network at the molecular level. The macroscopic evolution of organisms can also be visualized as the evolution of the basis set i.e. the gene/protein families and the interactions among their vector components. Thus, within the metric space of the connectionist sub-space of the protein similarity network universe, the protein families or clusters or partitions or modules of this sub-space form the dimensions of this metric space. The clusters or partitions that are equivalent to the dimensions of the metric space are approximate with weak dependencies between them (as will be subsequently recovered by the probabilistic graph clustering algorithm), esp. in the context of the stochastic nature of the evolutionary process that created this space. Gene families or protein families are one set of fundamental units of functional or basis sets that are functional dimensions of organisms or groups of species i.e. taxons. But the most fundamental unit of functional or basis set of dimensions is the domains which combine amongst themselves to create functionalities in general, and generate innovation in functionalities. The same is true of taxons which groups of organisms related by phylogenetic lineage. The study of the allocation of the subsets of the full basis set (of the entire living kingdom) in sets of organisms i.e. taxons is an important way to understand the similarities and differences among the taxonomic sets in terms of basis set of molecular functions. The organism or species or taxons are the units of systems which are composed of these unitary dimensions. The complexity of these functionalities and these system units or taxons, in one way, are best studied by the complexity of the functional dimensions composing them, the corresponding weightage of these dimensions in the taxons, which would give us the state vector of these dimensions by themselves and the systems units which they compose. The symmetries and complexities, the conservation and divergence of the dimensions themselves and the system units they compose gives the complexity, I one way of the dimensions themselves and the systems units they compose. The genomes of organisms, species or taxons or the system units encode this complexity. Now, deciphering the state vector, or more fundamentally their symmetry and complexity, their conservation and divergence gives an understanding of the complexity of evolutionary process itself and also fundamental design complexity of the solutions generated or created, in terms of functionality of the solutions (as for this paper work). The inferences, esp. from living kingdom and archaeal kingdom provide information about complexity of components responsible for connectionism, how levels of abstractions can be traversed and their complexity revealed which provide information about the nature and origin of connectionisms and the evolution of such complexity (computation capability or connectionist capability).

One major aspect characterizing complexity and its evolvability is the metric space or dimensionality of a sub-space. Within the field of molecular biocomplexity, this is valid for fundamental biological functional sub-spaces e.g. signaling and communication. Below we delineate the prime indicators of the molecular biocomplexity of the signaling system or any general biosystem or bioprocess, esp. when observed wrt the molecular functional (signal transduction) and organismal spaces of life (tree-of-life). For signaling (and communication), the holistic dimensions (protein family) and basis dimensions (domains) indicates how well connected sub-spaces with signaling dimensions are with their environment in the sense how much bandwidth of the spectrum of connectionism they interacts with and thus influences the totality (here environment). The ratio of ratio of functional members (protein nodes) to functional classes families indicates the density of sub-spaces corresponding to each of the dimensions of signaling, indicating the degree of evolution and/or evolvability of the dimensions themselves. Suitable annotations (e.g. Gene Ontology) can reveal the distributions of the functional classes of the dimensions. The ratio of functional classes or families to the total no of species reveal the span and degree of intersection of the observed and observer sub-spaces- a type of average complexity of signaling in a representative species. The distribution of the proteins nodes in individual species and their ranges giving the distribution of the density of one sub-spaces onto the other and the corresponding boundary conditions. The ratios (wrt total living kingdom) of signaling components or proteins in certain organisms reveal the relative node density of important sub-spaces i.e. species wrt the total sub-space under consideration. The distribution of the proteins in the three kingdoms of life and specific organisms/taxons from each kingdom esp. those having maximum and minimum numbers of the proteins per kingdom gives the structure of the observed tree-of-life sub-space and its major sub-spaces in terms of protein node density and the corresponding bounds- significant esp. for characterizing the evolvability of these species or taxons. The total no of unique domains in the entire living kingdom reveals the magnitude of spectrum of basis dimensions, in terms of proteins making up the functional aspect of the living kingdom. The distribution characterizing the no of domains present in numbers of protein (divided in ranges) gives the nature of the fundamental dimensionality of the protein nodes comprising the entire functional class (here signaling) in the living kingdom and the corresponding bounds, delineating the degree of generality of the domains and the distributions of degree of modularity and unit functionality in the nodes of the entire signaling sub-space. Distribution of copy number of domains and classification of the domains in terms of their copy number indicates the degree of degeneracy and the degree of modularity and multi-functionality of protein nodes that is of a particular class of degeneracy. This characterizes the relative roles of levels of degeneracy of unit dimensions of the metric space of signal transduction protein domains. This also characterizes the structure of the phase space of evolutionary dynamics of the unit functional dimensions i.e. protein domains. The range of protein cluster sizes and the no of clusters in that range signifies the fundamental density distribution of the space of protein families. The number of proteins, here signal transduction proteins and the relative ratios of those proteins in various important species characterizes the relative density of the species spaces within themselves in terms of proteins nodes of signaling space. The no of functional dimensions i.e. protein families in each species, and the classification of the species based on ranges of functional dimensions gives the complexity classes wrt signaling and communication of the species-space and the corresponding bounds on the complexity classes (given by the maximum and min number of protein families in various taxons) and ratios among them indicating relative complexity. The correlation between number of signaling proteins and number of signaling protein families or functional dimensions indicates the dependencies between the functional classes and density of protein components characterizing the phase space of evolutionary dynamics of the protein families. The sharing of dimensions among systems in metric space indicates the irreducible complexity of the space of systems containing those dimensions. Here sharing of signaling dimensions indicate how similarly environment can affect different systems. The kingdom-wise combination of species, the no of functional classes shared between these combinations, their ranges, no of species combinations having that degree of sharing; distribution of the degree of sharing and its max and min values characterizes the symmetry or similarity of signaling spectrum and its corresponding distribution with bounds indicates the symmetry or similarity of the functional dimensions within and across the sub-spaces of kingdoms of life. This indicates the symmetry and conservation of signaling protein families and the degrees of that conservation within and across kingdoms. The no of clusters or functional classes or families present atleast in a taxon gives the total dimensional of that taxon as a whole, characterizing the complexity and evolvability of the taxon as a whole. The distribution of these parameters across taxons gives the relative total complexity and evolvability of individual major taxons and characterizes the phase space of evolutionary dynamics of the total living kingdom. In the prokaryotic kingdom, since several taxons are known, separate characterization of such distribution is more significant in the prokaryotic kingdom. The pairwise (and other larger combinatorial) sharing of protein families across major taxons/kingdoms of the living kingdom indicates the total symmetry or similarity between these kingdoms in terms of the signal transductions. This would indicate the connectivity and their degrees and spectrums of each major taxon or kingdom with its environment and the same relative to other major taxons and kingdoms. The distribution of number of functional classes/protein families among degrees of sharing among taxons characterizes the nature of the evolutionary dynamics of the functional dimensions. The no of species/taxon combinations in different ranges of sharing of protein families signifies the general structure of the symmetry of evolution of the organismal space in terms of the common connectivities with their environment. The sharing capacity of each species with other species for signal transduction families, no of kingdoms with which shared, weightage of the kingdoms with which shared indicates the individual signaling symmetry of each species with the rest of the living kingdom and the distribution of this symmetry in the living kingdom for that species. For distributions of all species this characterizes some aspects of the complexity of the interactions within the living kingdom and the “complexity of symmetry” within the individuals of the organismal space. For every cluster or functional dimension, the number of kingdoms in which the functional class/dimension is distributed, no of species in each kingdom into which the functional dimension is distributed and degree of species-wise expansion/proliferation in each kingdom characterizes the distribution of the signaling dimensions or families in the living kingdom, upto the levels of individual species characterizing the penetration and its different levels of the individual dimensions, to better understand the previously states aspects of complexity. Exclusivity of taxons in terms of the signaling protein families and domains signifies the uniqueness of complexity and evolvability (in terms of holistic dimensions, unitary dimensions, relative weightage of unitary dimensions characterizing the holistic dimensional exclusivity of taxons, distributions of the degrees of unitary dimensions in the holistic dimensions) of taxons and taxons combinations and their exclusive evolutionary connectedness within the biocomplexity space. The domains present in different taxon combinations, i.e. either in individual taxons, or other total or exclusive taxon combinations characterizes the symmetry and complexity and evolvability of signaling functional metric space within and across the the ensemble of sub-spaces of the organismal space of tree of life.

The length ranges and distributions of the protein members are an indication of the information capacity (maximal informational capacity), being a measure of the information encoded and functional complexity of the members and relative diversity in terms of the informational complexity among themselves i.e. within the same functional class. Thus for a protein family, by itself, and when viewed wrt to unit systems like taxons, the length range and nature of variations within that range is a measure of the order of informational complexity and the process of evolution that generated that functional dimensions from the level of abstractions of the finite and close alphabet set. The number of species in which the families are present, species proliferation, indicates their spread across the living kingdom and thus their generality or specificity of utility within various organisms or taxons. The generality of the functional dimension within a set of species, given by the species spread gives the role of the strongly or weakly interacting dimension in the related system units, and indicates whether the dimension is specific to certain solutions or is general for a whole class of solutions and thus, also the generality of the computable capability of establishing connections of the species or taxons to their environment. The copy number of the families in the organisms in which they are present indicates the importance of the family in the organisms, their tendency for duplication within their host genomes and thus their volumes of requirement within the organisms (genomic dosage levels) and capacity to evolve further. Genomic dosage levels indicates the general strength and robustness of the computation capacity and also differentiation (e.g due to evolution) of resolution of functionality i.e. computation capability, indicating the sensitivity and degree of nonlinearity of the environment connectionists subspace with which such systems interact. The distribution of the functional dimensions, that is the weightage of the functional dimensions in each species/taxons gives the degree of importance of the role of that dimensions, whether it is a major factor or acts as weakly interacting factor in the systems units. The no of different domain architectures making up the family indicates the complexity of the families evolutionary history and its tendency to diversify in structure and function, the differentiation and resolution of such functionality achieved and thus enabling the variety of interactions with environment. The length of the domain architectures and their distribution among members indicates the degree of modularity of the family members and relative distribution among equi-functional set, the complexity of the interactions as reflected by the number of individual basis dimensions required and in what weightage the interactions are present. The frequency distributions of the domain architectures within the families indicates the presence of any preference for any architecture among the family members and thus any bias towards a particular architecture or form of modularity, indicating any relative importance of those functional differentiated dimensions and thus any specific bias for any kind of dynamics with the environment within the connectionist paradigm. Their weightage of distributions within the family gives the idea how well-bounded that functional dimension has been in the process of evolution implying the stability of the dimension and its nature of evolution. The no of unique domains making up the architectures of the family indicating the potential or degree of multi-functionality amongst the members and thus their functional complexity, - a prime factor in signaling and communication systems, reflecting the fundamental complexity of these computational components or dimensions, which is equivalent to the complexity of the fundamental interaction within the connectionist totality. The frequency distribution of the unique domains among the family members indicates any bias for any or some particular domains and thus the strongest basis for their homology/common ancestry and functionality. It also indicates the relative weightage of the unitary dimensions within the holistic dimensions- which establishes the strength of unity and fundamental symmetry within a complete holistic dimension i.e a protein family and the nature of that symmetry. It gives the complexity of those dimensions from the levels of abstraction of the unitary functionalities, and is a different complexity from that available from the length of amino acid sequence. It indicates the multidimensionality, multi-modularity and multi-functionality of the holistic dimensions of protein families. The no of unique domains also indicates the potential or actual degrees of connectivities form another observer sub-space. The actual values of the above parameters and dependencies or relationships within and across these parameters, within and across living kingdoms will provide a complexity map. From a certain frame of reference they do give a complete picture of the complexity of the protein families of a particular generalized biological functionality. This certain reference frame can be considered to be the structural nature of the members of the families and their basic tendency for abundancy in systems in which they are present i.e. existence and their distributions.

The archaeal kingdom is uniquely different from the other two kingdoms. This extraordinary and intriguing uniqueness of the archaea immediately compels us to focus on its signaling system, to see what kind it is, how much it is similar to that of the other two kingdoms of bacteria and eukaryotes and how much it is different from them, how much the dependencies are similar and different within and across the kingdom or lineage specific protein families, within and across kingdoms and lineages. This requires the creation of a complexity map of the archaeal signaling components i.e. i.e. the identification and delineation of the variations and dependencies or correlations within and across the archaeal signal transduction protein families in terms of the fundamental parameters determining its complexity i.e. their biological names and functions, family sizes, information capacity and functional complexity (lengths of protein members), their taxonomic spread and copy numbers, their variety of domain architectures and their length and frequency distribution among the members, their modularities and multifunctionality i.e. component domains and their usage and frequency distribution.

One of the purposes of the complexity-theoretic analysis was to discover new or appropriate questions that can or should be asked. Since any significant complex sub-space is colossally complex, the diversity of the sub-space is usually enormous. Thus the classification of diversity is extremely difficult and it becomes impossible to directly map complexity theoretic inferences of real phenomenon directly to universal ontology of complexity. This is also due to the language of the phenomenon, here biology, which is quite far off from complexity theoretic ontology. Also, since the objective is to study the complexity of colossally complex sub-spaces, the possible spectrum of possible inferences is enormous. So it is practically impossible to delineate spectrum of possible inferences and subsequently map the biological (here) inferences directly with the universal ontology. Thus to delineate the complexity of complex sub-spaces according to the theory of complexity provided in the paper, a case by case study has to be performed to provide complexity theoretic inferences. This possibly generates a significant spectrum of the inferences But still that is inadequate to map the inferences to unambiguous set of complexity theoretic inferences i.e. the universal ontology of complexity. But work in this paper significantly moves in that direction through the study of signaling and communication- themselves complex and responsible for complexity in living kingdom, which is also colossally complex. This form of exercise provides a spectrum of complexity theoretic inferences on real systems and thus significantly showing the path for universal ontology of complexity through real and valid case studies. Thus, below we discuss the major results obtained from the complexity theoretic analysis of signaling in living kingdom, from within the philosophical perspective discussed to this point. Ofcourse, as more and more data will be included some inferences may and will changes, but some will remain the same esp. those of more general nature. But, all the inferences significantly portray the possible ontology of complexity, which is more important than the exact sustenance of the specific results, say for a specific set of data, over time.

Conclusion

The primary philosophical theme of this paper is the concept of Complexity. The work explores the concepts of complex systems, processes and phenomena from the complexity theoretic perspective, incorporating the connectionist conjecture, the concepts of signaling and communication in complex systems, the concept of spaces and sub-spaces, their metric spaces and dimensionality. The paper attempts to integrate these concepts on a representative true complex system or an ensemble of complex systems i.e. biological systems and biological functions - their complexity theoretic analysis leads to a better understanding of complexity in general and specifically of biocomplexity. Since biological systems are naturally engineered, they have undergone a long range dynamics representative of an evolutionary multi-objective optimization, in terms of their interaction and stability with their environmental spectrum, where the inherent design, structure and dynamics of such systems have an adaptive engineering, with abilities of self-organization, criticality and emergence, with a pervasive embedded randomness. The paper emphasizes the concepts of information and non-linearity, either in structure or in dynamics or both, in such multi-level complex systems, as represented by their states, where the patterns in distributions and fluctuations of states are prime indicators of their complexity. Availability of large-scale measurements of the states of these complex systems, organized in complex knowledge-bases, provides scope for such complexity theoretic analysis.

The following summarizes the contributions of the paper.

- The basic objective of this paper was to forward ideas and concepts on complexity towards new world views, subsequently integrate the real world e.g. the living world within that world view, present and discuss ways of approaching and studying the real world e.g. living world, in the new world view and, attempt analysis of aspects of the real world e.g. biomolecular function and taxons in the living kingdom, and to demonstrate the form of knowledge gained from that new world view.

- The paper treats the evolution of genes, domains, genomes and gene families, evolution of biological functions, the basic nature and complexity of signaling and communication in living world, protein domains/modules or unitary functional dimensions, nature of major taxonomic sets and their inter-relationships in terms of the holistic and unitary dimensions - all with a complexity theoretic flavor. The paper treats taxa from tree of life as bio-systems and signal transduction families and modules or domains as bio-functionals, and several distributions and corresponding ranges and fluctuations as qualitative measures of patterns, symmetry and complexity.

- For analysis i.e. a) characterizing the complexity of major taxonomic sets and their interrelationships in terms of signaling protein families/holistic signaling functional dimensions and domains/unitary functional dimensions and b) characterizing the complexity of one exotic kingdom of life, archaea, esp. its uniqueness and commonality or symmetry with the rest of the living kingdom, all the parameters (mentioned in above two points) and their distributions studied, and corresponding inferences derived have been specifically attempted to be interpreted within a narrow set (dimensionality, symmetry, information, nonlinearity, degeneracy etc.) of complexity theoretic concepts and corresponding terminology, to converge towards an Ontology of Complexity for higher order integrative meta-analysis of complexity and complex systems.

- **The main contribution of this paper is the Connectionist Conjecture** which proposes a world-view or nature of totality as the connectionist representation i.e. a complex network- a graph with nodes and edges, being the one and only representation of totality and its sub-spaces.

- The paper forwards an array of ideas and concepts revolving around the connectionist conjecture - irreducible complexity of totality, inherent randomness in totality, recursive and perpetual evolution of information, computability of totality and its sub-spaces, conservation and symmetry, equivalence of symmetry and complexity, emergence of criticality, ensemble of sub-spaces, metric spaces and dimensionality of sub-spaces, states, their distributions, ranges and fluctuations, and corresponding patterns. It focuses on the role of signaling and communication in evolution and maintenance of the connectionist structure of totality of the living world or its sub-spaces, and possible complexity theoretic interpretations of molecular biological parameters like protein families, domains or modules, domain architectures, genomic dosage levels, proliferation, functional differentiation, dynamical nature and robustness of parameters, and total & exclusive taxon combinations.

- The paper advocates that one way of characterizing and understanding systems, processes or phenomena is to study them in an no-a priori-hypothesis based holistic approach, where a kind of “matrix” of inferences and their dependencies are created, to characterize the whole entity, without restricting to a (generally) narrow set of inferences constrained by one or more a priori hypothesis (as is done most of the times). The paper strives to demonstrate that, although the complete representation of the entity (under study) may not be normalized (i.e. having a uniform representation at all levels of abstractions and projections e.g. a complex network with node and edges and numerical distributions of properties or parameters embedded in them) but rather having a heterogeneous representation across levels of abstractions and projections (usually true for biological phenomena), this ab initio no a priori hypothesis based holistic approach yields inferences (a significantly large number, when true complex systems are considered), a large part of which are generally useful and also has the potential to raise further questions that would have been missed/overlooked in the popular a priori hypothesis constrained studies. The paper advocates that this is an extremely important issue for the study of complex systems, process and phenomenon, in general but also specifically in the context of their connectionist representation, where the cause-effect relationships are blurred, insignificant or even meaningless.

In all, the Connectionist Conjecture and the associated ideas esp. the Ontology of Complexity, embodied in this paper, including the extensive analysis of molecular biocomplexity, can play an important role for understanding the Complexity of Totality.

References:

1. “Explorations on Complexity and Signaling- Computational Analysis of Compositional, Structural and Functional Features within and across Genes and Genomes”, Debaprasad Mukherjee, PhD Thesis, Faculty of Engineering and Technology (FET), Jadavpur University, Kolkata, India (submitted May 2009)
