Informatics for Discovering the "Inorganic Gene" in Complex Inorganic Solids

Krishna Rajan
Department of Materials Science and Engineering
Iowa State University
Ames, IA. 50011

Through many seminal papers, crystallographer Alan McKay has expounded on the idea of a framework for "Generalized Crystallography". He has proposed that the crystal is a "structure the description of which is much smaller than the structure itself" and that this description of structure serves as a "carrier of information" about the structure on larger length scales. He went on to suggest that these components of description of structure can help develop a "biological approach to inorganic systems" and proposed the construction of an "inorganic gene". This paradigm serves as motivation underlying the present study by exploring how fundamental pieces of information, treated as discrete bits of data, can collectively characterize the stability and properties of a given crystal chemistry. We show how the use of statistical learning tools and information theory can be used to characterize a crystal structure in terms of fundamental descriptors of information (ie. the "genes") and how these pieces of information interact or are "sequenced" to guide the characteristics of that crystal structure. It is suggested that this type of analysis could in fact help to guide the development of the next generation of key experiments using hard X-rays.