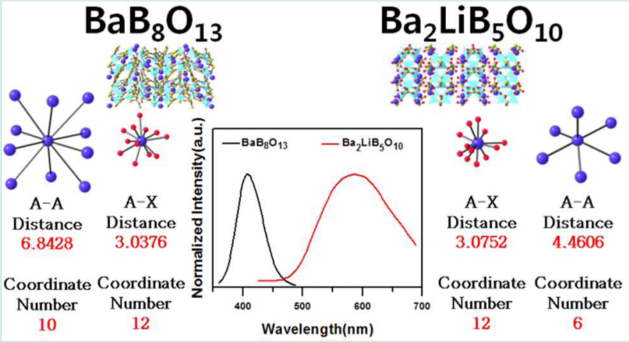
Phosphor Informatics Based on Confirmatory Factor Analysis

Woon Bae Park, Satendra Pal Singh, Minseuk Kim, and Kee-Sun Sohn*

Faculty of Nanotechnology and Advanced Materials Engineering, Sejong University, Seoul 143-747, Korea

S *Supporting Information*

ABSTRACT: The theoretical understanding of phosphor luminescence is far from complete. To accomplish a full understanding of phosphor luminescence, the data mining of existing experimental data should receive equal consideration along with theoretical approaches. We mined the crystallographic and luminescence data of 75 reported Eu^{2+}-doped phosphors with a single Wyckoff site for Eu^{2+} activator accommodation, and 32 descriptors were extracted. A confirmatory factor analysis (CFA) based on a structural equation model (SEM) was employed since it has been helpful in understanding complex problems in social sciences and in bioinformatics. This first attempt at applying CFA to the data mining of engineering materials provided a better understanding of the structural and luminescent-property relationships for LED phosphors than what we have learnt so far from the conventional theoretical approaches.



KEYWORDS: *materials informatics, materials genome, data mining, light emitting diode, solid-state lighting, phosphor, combinatorial materials search, principal component analysis, confirmatory factor analysis, structural equation modeling*

1. INTRODUCTION

Enormous effort has been invested in establishing either a theoretical or empirical model that will enable an accurate prediction of the luminescence of phosphors. Of particular interest is a reliable prediction of emission wavelength, emission bandwidth, and quantum efficiency, which has been a challenge in association with the quest to discover novel phosphors for use in light emitting diodes (LEDs). The ab initio approach based on density functional theory (DFT) calculation has consistently improved, and many structural and physical properties have been accurately evaluated.[1] As far as the luminescent materials are concerned, however, a reliable calculation-based prediction is yet to be accomplished. The DFT calculation has only recently reached the level of a rough estimate of the ground state of $4f^n$ and $4f^{n-1}5d^1$ energy states for lanthanide activators, albeit with no precision.[2−4] It has been impossible to precisely predict emission wavelength, emission bandwidth, and quantum efficiency with satisfactory accuracy, no matter what types of currently available theoretical approaches have been adopted.

In contrast to the current status of the calculation of luminescent materials, however, Ceder et al.[5−7] achieved a remarkable advancement by combining theoretical ab initio calculation with a heuristic data mining approach and thereby a prediction of the feasibility of suggested structures, as well as a prediction of material properties, has been achieved, particularly for binary intermetallics (or alloys) and some Li-battery materials. However, such brilliant achievement has never been transferred to the study of phosphor materials because typical LED phosphors are more complex, with a greater number of constituent elements (e.g., quinary, senary, or more), and novel phosphors discovered outside the prototype structures are more desirable due to the intellectual property (IP) complication in the field.[8,9]

Analogous to such a heuristic data mining approach, we have employed the so-called metaheuristics-assisted combinatorial materials discovery strategy, which has resulted in several outstanding, practical discoveries of phosphors for use in LED applications.[8−10] This approach was more oriented to experiments based on high-throughput syntheses, and therefore the main focus was a substantial discovery rather than a prediction. This approach enabled us to discover novel inorganic compounds, which never belonged to any of the well-known prototype structures, which the Inorganic Crystal Structure Database (ICSD) classified based on reasonable criterion.[11] A total of 8230 prototypes were announced this year.[12] The metaheuristics-assisted combinatorial materials discovery strategy has led to the discovery of novel prototype structures. A huge amount of experimental data was produced during this discovery process. However, the data was based only on stochastic choice with no in-depth data mining or the ensuing understanding.

The main focus of the present investigation originated from the statistical approach that is typically used in the field of social sciences. We introduced confirmatory factor analysis (CFA)[13,14] for data mining of existing phosphors by setting

Received: January 29, 2015
Revised: April 7, 2015
Published: April 8, 2015

DOI: 10.1021/acscombsci.5b00017

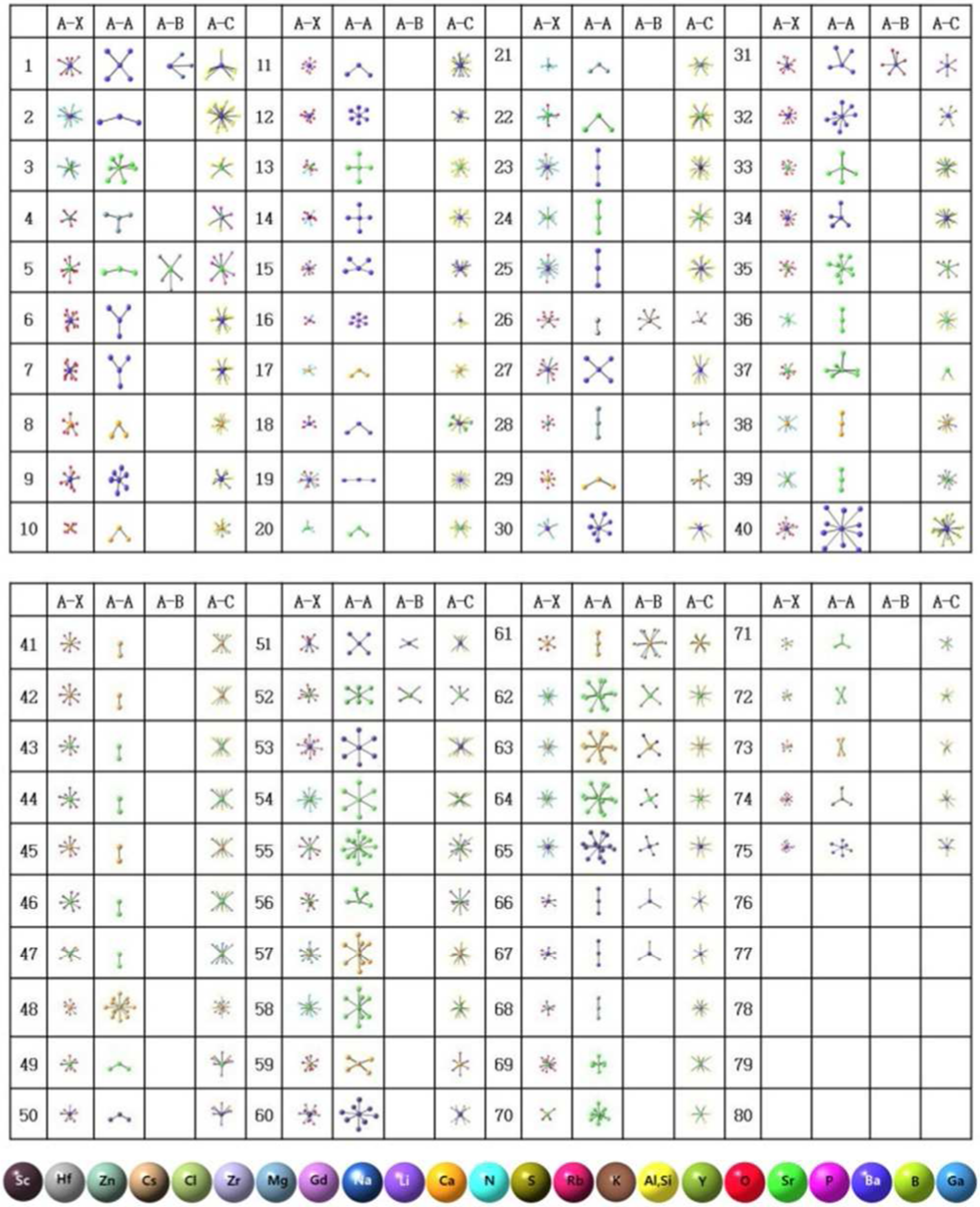


Figure 1. A–X, A–A, A–B, and A–C local structures. Atoms are represented by color, as shown in below. Also, the relative length obeys the actual length scale. The number represents corresponding phosphors listed in Table s1 in the Supporting Information.

up the most appropriate structural equation model (SEM), which provided us with a useful guideline for the design of novel phosphor discovery. CFA differs from conventional regression (or machine learning) technique, because CFA focuses more on understanding rather than regression—a complete understanding of the underlying relationship between latent factors through an advanced SEM.

A number of pioneering achievements have been reported in similar frameworks, wherein principal component analysis (PCA), partial least-squares (PLS) regression, and some other machine-learning techniques, such as artificial neural network (ANN) and support vector machine (SVM), have been employed for the data mining of other material systems.[15−22] In particular, Rajan et al.[22] have recently reported interesting results by data-mining perovskite, apatite structures, and some metals. Our approach, however, had two distinct points that contrasted with these conventional data mining cases. First, the material that we adopted for data mining was luminescent materials (phosphors), which should be the most suitable material for data mining because the complexity of the material leads to incalculability. Second, a more significant distinction was the uniqueness of the statistical approach that we adopted. CFA has never been used for a materials data mining process, although it has proven to be a powerful tool that can promote a better understanding of high dimensional data sets, particularly in the social sciences and bioinformatics fields.[13,14,23,24]

2. DATA ACQUISITION AND DESCRIPTOR EXTRACTION

The data acquisition was the most significant part of the meta-analysis. From the literature, we collected structural and luminescent data that are listed in Table s1 in the Supporting Information. For the sake of accurate data acquisition, we downsized the scope of the material, first by restricting it to Eu^{2+}-doped, oxide, oxyhalide, oxynitride, and nitride phosphors. The 5d energy of Eu^{2+} and Ce^{3+} in oxide, oxyhalide, oxynitride, and nitride phosphors was scattered widely.[25,26] Thousands of Eu^{2+}-doped phosphors have been reported thus far, but we restricted our scope to those having only a Wyckoff site for activator accommodation in the host structure. Phosphors with two or more activator sites presented complications in determining the emission peak wavelengths and widths because of the unclear activator site assignments and the energy transfers between different activator sites. Finally, we precluded cation solid-solution types from the analysis. Consequently, we secured 75 phosphors for the statistical analysis.

We screened every reported Eu^{2+}-activated phosphor by referring to Scopus databases, and to the inorganic compound structure database (ICSD). As a result, we found more than 2500 papers regarding the Eu^{2+}-activated phosphor. The number of Eu^{2+}-activated phosphors reported thus far, which exhibit distinctive crystallographic structures, was a total of 258. Of those, 139 Eu^{2+}-activated phosphors had a single Wyckoff site for an Eu^{2+}activator, and 119 had more than two. We further reduced the number of Eu^{2+}-activated phosphors by excluding some old-fashioned, less-stable phosphors based on sulfides and halides. Thus, we adopted 83 single-activator-sites, and Eu^{2+}-activated phosphors based on oxides, oxyhalides, nitrides, and oxynitrides, which were targeted for use in LED applications. Finally, a series of cation solid-solution phosphors based on the same structure was removed, and the resultant number of phosphors adopted in the present investigation was 75.

The aim of the present investigation was to understand the luminescence of Eu^{2+}-activated phosphors using their structural characteristics and the basic characteristics of their constituent elements. We denoted the structural and elemental characteristics of phosphors as "descriptors". As for the material descriptor extraction, 32 material traits were selected for use in our statistical data mining approach. Most of the descriptors were related to the crystal structures of the host materials. We did not involve thermodynamic parameters and physical properties of host materials as a descriptor. It is noted that traditional luminescence study has never focused on thermodynamic parameters. Some physical properties, such as the dielectric constant (refractive index) of host materials, would be closely related with luminescence but it is substantially impossible to obtain them for the 75 complex material systems. The latent factors in SEM might incorporate these missing descriptors although we do not know how they worked in detail.

The chosen Eu^{2+}-activated phosphors can be expressed by the ANX formula[12] for systematic descriptor extraction as

$$A_aB_bC_cX_x\text{: }Eu^{2+}$$

where A, B, and C stand for the cation sites, each of which has an independent Wyckoff site, likewise, X stands for anion sites. A is the activator site, normally the alkali-earth-element site in the host compound; the B site denotes the nonactivator cation site. Many phosphors had no B site in their structure. The C-site element was the most important in determining the entire structure of a host, because these created either a tetrahedron or an octahedron network along with anions. Such a network creates a structure with either a two- or three-dimensional backbone. The network forms by corner sharing through either bridging or tripling points. The C-site element normally is a light, small element such as Si, P, B, or Al; sometimes Li, Sc, or Mg may occupy the C site. The small characters, *a*, *b*, *c*, and *x* denote the stoichiometry of the host compound.

The basic idea to extract structural descriptors was the reconstruction (or reinterpretation) of the host crystal structure by centering on activator sites. Both the activator-anion ligand polyhedron and the local structure comprised of the activator and one of the other elements were parametrized as descriptors. Only the activator site and its first anion neighbors were the focus in most of the previous investigations. On the contrary, we involved A−A, A−B, and A−C polyhedra consisting of the nearest neighbors around the A site. In this regard, the coordination number and the average distance of every polyhedron around the activator site were defined as descriptors, such as C_{A-X}, C_{A-A}, C_{A-B}, and C_{A-C} for the coordination number, and d_{A-X}, d_{A-A}, d_{A-B}, and d_{A-C} for the average distance. In the actual statistical process, we used the reciprocal value of the distance because we had to reasonably account for phosphors with no B ion site in the structure, such that $1/d_{A\text{-}B}$ equaled zero if a B site was nonexistent.

The local structures around the Eu^{2+} activators for 75 different hosts are presented in Figure 1. Each polyhedron was made up of the nearest neighbors consisting of X, A, B, and C elements, respectively. The nearest neighbor was determined at the first substantial rise in the magnitude of interatomic distance. However, in a few cases where the interatomic distance continuously increased, the nearest neighbor was at a distance that had changed by more than 10%. Some of the A−A and A−B local structures were not a polyhedron, but resulted instead in 1- or 2-dimensional shapes. These nonpolyhedron type local structures could be also meaningful, because these must have been acting as interactivator energy transfer routes, which will be discussed in more detail in the section 4. The activator-anion local structure was a major concern in previous investigations, and no attempts were made to examine the cation−cation local structure around the activator site. The cation neighbors should have a significant degree of influence on the luminescence, but the mechanisms are unknown.

In addition to the eight structural descriptors, another 12 descriptors indicating the constituent elements occupying the A, B, C, and X sites were defined. For instance, the atomic number (N_A, N_B, N_C, and N_X), the electronegativity (E_A, E_B, E_C, and E_X) of every constituent atom in the Pauling scale,[27] and the Shannon radius (R_A, R_B, R_C, and R_X) in the corresponding local environments[28] should affect the bond character. In evaluating these elemental characteristic parameters, a special consideration was made when more than one element constituted each of the A, B, C, and X sites. In these cases, weighted average values were obtained according to the relative number of atoms constituting each polyhedron. Lattice parameter anisotropy, lattice angles, lattice volume, and theoretical density were also adopted as descriptors (b/a, c/a, β, γ, V and ρ). Because there was no triclinic in our data set, α was omitted. Finally, we adopted basic symmetry descriptors such as the space group number (SG), the activator site symmetry number (SS), and the activator site multiplicity (AM). All these structural and elemental descriptors were defined such that the phosphor host structure could be reinterpreted by centering on the activator site.

In addition to both the structural and elemental descriptors, three luminescence-related descriptors were also added. The emission peak wavelength (PW) and the full width at half-maximum (fwhm) were adopted as descriptors in units of eV along with the critical Eu^{2+} activator concentration (x_c) in units of cm^{-3}. PW and fwhm were obtained at x_c. In fact, x_c was the Eu^{2+} activator concentration that exhibited the highest PL intensity. The emission peak wavelength is known to vary dramatically with the Eu^{2+} activator concentration, so that the peak wavelength should never be regarded as a material's intrinsic property. However, the concentration-quenching data, that is, the emission spectra data that were monitored as a function of the Eu^{2+} activator concentration, were available only for 45 out of 75 phosphors. When the concentration-quenching data were not available, the Eu^{2+} activator concentration, for which PW and fwhm values were evaluated, was approximated to x_c. It is reasonable to assume that the best samples of PL intensity were used in most of the recent LED phosphor-related reports, unless the measurement of the radiative decay time was attempted. To ensure this assumption, we omitted such reports that deals with the detailed spectroscopy in diluted model phosphors at cryogenic temperatures, when the data set was acquired.

If it was possible to obtain zero phonon energy value as a luminescence-related descriptor, the data mining would be more promising. In such an ideal case, the data mining would be unnecessary because a theoretical approach could suffice. However, it is practically impossible to collect the zero phonon line data for every phosphor. The zero phonon line data collected for extremely diluted model phosphors at cryogenic temperatures could be more attractive than what we adopted as luminescence-related descriptors. However, those data are extremely scarce. The conventional emission data measured at room temperatures for practically acceptable activator concentrations were only available in the field. It should be noted that the data mining is always based on the general data with a certain degree of errors. The ultimate goal of the data mining is to acquire useful information from such a limited, incomplete data set. It should be noted that the practical data treated in the physical science are also incomplete and erroneous. It is conventional to systematically treat the error as a parameter in most of statistics-involved data mining processes. Thus, the incomplete x_c data would not be problematic, as far as the statistical science was of concern.

Table s1 in the Supporting Information shows the chosen 32 descriptors and their evaluation results for 75 different Eu^{2+}-activated, single-A-site phosphors. The 32 descriptors are divided into three categories, the host structure descriptors, the constituent element descriptors, and the luminescence descriptors. Both the host structure descriptors and the constituent element descriptors were regarded as indicator (or predictor) variables, whereas the luminescence descriptors became target variables. In particular, it should be noted that the activator concentration dealt with here is the critical activator concentration (x_c), not an indicator variable but a target variable, such as PW and fwhm.

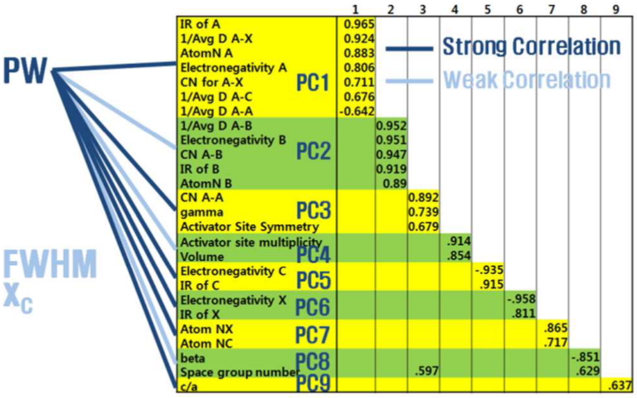
3. PHOSPHOR INFORMATICS THROUGH CONFIRMATORY FACTOR ANALYSIS

Confirmatory factor analysis (CFA)[13,14] was employed for a statistical approach to a plausible interpretation of the descriptor data collected for Eu^{2+}-activated phosphors with a single activator site. In general, CFA differs from conventional regression because it involves a set of latent variables in identifying the underlying relationship between measured variables. In addition to the indicator variables introduced in the section 2, we incorporated several latent variables via principal component analysis (PCA) and the ensuing linear regression using several principal components. The preliminary PCA and regression process was for data dimensionality reduction as well as for the rough determination of latent factors in advance of the CFA. The latent variables (or factors) should implicate some of the material attributes, but it is impossible to measure them. The aim of CFA was not to develop a regression model that could predict emission properties as a function of the host structure and the constituent element. The empirical regression model is nothing but an inattentive attempt, which usually is futile even though there have been a number of such research attempts in the field. Our aim was to more systematically understand the complicated correlations among the descriptors.

The conventional approach precludes target variables in the PCA. Thus, PCA was implemented for 29 structural and elemental descriptors. The 75 phosphors were distributed in a 29-dimension hyperspace, where each basis axis represented 29 descriptors. PCA enabled us to reduce the data dimension enough to be understood by human cognitive ability, while minimizing the loss of information during this data reduction. In fact, PCA was a part of CFA, so that it was carried out prior to conducting the CFA process to choose appropriate indicators, which could indirectly (or partially) measure each latent factor. Consequently, we selected 9 principal components (PCs) with Eigen values greater than 1. Indicators for each of the PCs were determined such that their factor loading exceeded 0.5. This criterion is usually adopted in the field of social sciences.[13,14] Thereafter, a simple linear regression was implemented using the 9 PCs as predictors and the emission peak wavelength (PW) as a target variable to roughly check the correlation between the target variable and each PC. The regression was not available when the full width at half-maximum (fwhm) and the critical activator concentration (x_c) were adopted as a target variable. Only the PW was well fitted to the regression model. As a result, a certain degree of correlation was found between the PW and the 6 PCs (PC1, PC3, PC5, PC6, PC7, and PC9). The PCA and the ensuing regression results are presented in Table 1. The result was used as the baseline for a reliable specification of the structural equation model (SEM) for the ensuing CFA.

As for specification of the SEM in the CFA, the overall performance factor (or target factor) determination was the

Table 1. PCA and the Ensuing Linear Regression Result[a]



PW
FWHM
x_c
Strong Correlation
Weak Correlation

		1	2	3	4	5	6	7	8	9
IR of A	PC1	0.965								
1/Avg D A-X		0.924								
AtomN A		0.883								
Electronegativity A		0.806								
CN for A-X		0.711								
1/Avg D A-C		0.676								
1/Avg D A-A		-0.642								
1/Avg D A-B	PC2		0.952							
Electronegativity B			0.951							
CN A-B			0.947							
IR of B			0.919							
AtomN B			0.89							
CN A-A	PC3			0.892						
gamma				0.739						
Activator Site Symmetry				0.679						
Activator site multiplicity	PC4				.914					
Volume					.854					
Electronegativity C	PC5					-.935				
IR of C						.915				
Electronegativity X	PC6						-.958			
IR of X							.811			
Atom NX	PC7							.865		
Atom NC								.717		
beta	PC8								-.851	
Space group number				.597					.629	
c/a	PC9									.637

[a]Loadings to each principal component are listed (VARIMAX rotation was adopted). The emission peak wavelength (PW) was a target variable with nine principal components as predictors. The color of the line represents the degree of correlation; the dark line represents a strong correlation and the dim line shows a weak correlation. FWHM and x_c had no acceptable correlation with any of the nine principal components.

first step. PW, fwhm, and x_c were used to measure the target factors. However, no other factors could be connected to the performance factor with sufficient statistical significance. Therefore, the performance factor was removed from the SEM and instead only a PW indicator was used as a target variable to be identified. Similar to prior regression, the inclusion of fwhm and x_c deteriorated the statistical significance. The only target variable adopted in the present approach was PW.

The final specification of the SEM was set up after testing a large number of plausible specifications using the PCA results, as shown Figure 2. As a result, PC2, PC4, and PC8 were removed in the final model since they showed no correlation with PW in the linear regression and neither statistical significance nor intuitive accountability was obtained when they were introduced as factors in the SEM. Although PC3, PC7, and PC9 showed a correlation with the PW in the linear regression, it also deteriorated the statistical significance and intuitive accountability as factors in the SEM. Accordingly, they were also removed based on the trade-off between parsimony and goodness-of-fit.[29]

PC1 consisted of local environmental indicators around the activator site, such as A−X, A−A, and A−C distances, as well as A-site element descriptors. However, when PC1 was set as a latent factor with its indicators, the statistical significance was never obtained at the 0.05 level. So PC1 should never be used as a latent factor, but Instead, PC1 was split into four factors, each of which represented the A−X, A−A, and A−C local environments, and an additional factor representing the A element traits. Thereafter, all the others were removed except for the A−X and A−A local environment factors, because the removed factors also led to an unacceptable level of statistical significance.

All of the latent factors basically came from the PCs with a certain degree of correlation with PW; namely, two latent factors directly from PC5 and PC6, and the other two from the PC1 splitting and also from PC3. We tried every possible SEM using these four factors. In particular, an effort was made to link each factor either directly or indirectly to the target variable (PW), and the final specification was determined based on the statistical significance, which should be far below the 0.05 level. There was a clear causal effect of the A-A local environment factor on the A-X local environment factor. It should be noted that other types of data mining strategies, such as linear or nonlinear model regression, artificial neural network, support vector machine, etc.,[15−22] could never identify this type of indirect effect. Figure 2 shows the final SEM, and the standardized regression weights for every factor and indicator are summarized in Table s2, which is available in the Supporting Information.

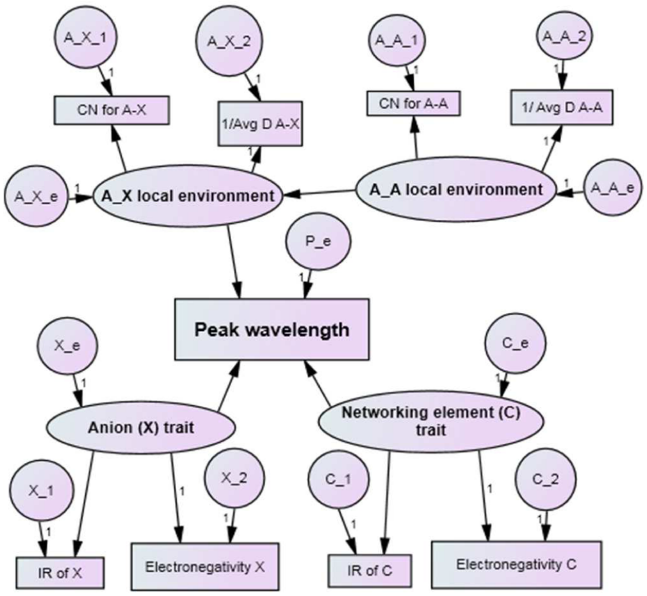


Figure 2. Structural equation model (SEM) involved four latent factors: The ovals are the A−X local environment factor (or small scale environment factor), the A−A local environment factor (or large scale environment factor), the anion trait factor, and the networking element trait factor. The rectangles are indicators. The circles are error terms. The number "1" indicates that the variance is fixed as 1.

When the final specification of the SEM had been completed, as shown in Figure 2, a plausible title was required for each factor. The title for a factor was determined such that the overall meaning of the indicators measuring the factor could be incorporated in the title. The first two factors were titled A−X and A−A local environment factors. The indicators measuring these factors were d_{A-X}, C_{A-X}, d_{A-A}, and C_{A-A}. Both d_{A-X} and C_{A-X} described the smallest anion polyhedron around the A site, which is well-known to account for the 5d energy of Eu^{2+} (or Ce^{3+}), and has played a crucial role in Dorenbos model.[25,26] d_{A-A} and C_{A-A} represented the activator site distribution centering on the activator site (A site). Although traditional phosphor research has never focused on the activator site distribution around the activator site, d_{A-A} and C_{A-A} played a certain role in our SEM. The A site atom was normally one of alkali earth elements, and its choice was considered to have affected the anion polarizability, which therefore had a slight influence on the 5d energy shift.[25,26] However, the A−A distance and coordination number has never been of interest. The detailed A site consideration was unprecedented in the previous study since the accepted wisdom dictated focusing only on the anion polyhedron around the activator site.

Finally, the factors from PC5 and PC6 were titled the anion trait factor and the networking element trait factor, respectively, since PC5 and PC6 represented the attributes of the X and C site atoms, respectively. The indicators measuring these elemental factors included the ionic radius (R_X and R_C) and the electronegativity (E_x and E_C). Both the anion trait factor and the networking element trait factor directly affected PW with no causal effect on other factors.

The four latent factors were titled the A−X local environment factor (or small scale environment factor), the A−A local environment factor (or large scale environment factor), the anion trait factor, and the networking element trait factor. The introduction of the A−X local environment factor and the anion trait factor in the SEM was not surprising, since common sense suggests that the A−X local environment plays a key role both in the crystal field theory and in the nephelauxetic effect theory. Therefore, PW depends on the A−X distance and more importantly on the anion type, such as fluoride, chloride, sulfide, oxide, or nitride. The networking element trait factor was also reasonable because PW was greatly affected by the type of networking atom, since aluminate, silicate, borate, phosphate, etc., are distinctive. This shows how the final specification of SEM was in complete agreement with the traditional scientific finding.

Even more interesting was that the A−A local environment factor had an indirect influence on PW through the A−X local environment factor. This implies that, relative to PW, the A−X and A−A local environment factors had a strong correlation. This reflected a causal effect of the A−A local environment factor on the A−X local environment factor in the final SEM. The effective local environment around the activator spanned beyond the anion local structure and thereby the cation distribution around the activator should be considered when planning for the discovery of novel phosphors. It is not surprising to see such a causal effect because the A−A and A−X local structure are interconnected in the host structure. For example, a distortion in A−A polyhedrons would definitely lead to a certain degree of change in A−X polyhedrons. This clearly indicates that the SEM derived from the experimental data set was reasonably explained by the scientific finding, and simultaneously provided better understanding in comparison to the theoretical interpretation only.

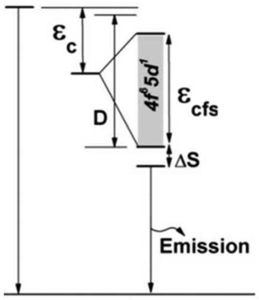
Although the A−A local environment has never been of interest in the conventional phosphor research focusing on the crystal field and covalency effects only, the role of the A−A local environment factor could be well understood on the theoretical basis. Besides the connected distortion in the A−A and A−X polyhedrons, the most reasonable interpretation for the A−A local environment factor would be related to the interactivator energy transfer. It is obvious that the interactivator energy transfer should have a great influence on both the peak wavelength and fwhm. The energy transfer route characterized by the A−A distance and coordination number plays a critical role in determining the energy transfer rate.[30−33] In this regard, the inclusion of the A−A local environment factor in the final SEM was reasonable form the theoretical point of view.

The measured PW could be split into intrinsic and extrinsic terms (PW = $\mathrm{PW}_{\mathrm{int}}$ + $\Delta\mathrm{PW}_{\mathrm{ext}}$) to give a theoretically reliable interpretation to the final SEM. What we can infer from the typical Dorenbos model[25,26,34,35] is $\mathrm{PW}_{\mathrm{int}}$, which can be measured at a very diluted system based the assumption of inactive interactivator energy transfer, while $\Delta\mathrm{PW}_{\mathrm{ext}}$ is dependent on the activator concentration. $\Delta\mathrm{PW}_{\mathrm{ext}}$ is greatly affected by the interactivator energy transfer and partly by the host lattice distortion. In this context, the A−X local environment, the anion trait factor, and the networking element trait factor played a significant role in determining $\mathrm{PW}_{\mathrm{int}}$, while the A−A local environment factor was closely associated with $\Delta\mathrm{PW}_{\mathrm{ext}}$.

The details of the final SEM will be discussed in the following section, using a specific case wherein conventional theory could not provide a reliable interpretation, but the present SEM could. Two very similar phosphors, $\mathrm{Ba_2LiB_5O_{10}:Eu^{2+}}$ and $\mathrm{BaB_8O_{13}:Eu^{2+}}$, exhibited totally different emission colors, with different A−A environments and a slight C-site atom alteration. Although the A−X local structures were similar, the phosphors could be differentiated based on the SEM.

4. UNDERSTANDING OF THE STRUCTURAL EQUATION MODEL USING A SPECIFIC CASE

The motivation of the present investigation was the brilliant work of Dorenbos,[25,26,34,35] wherein a very simple, reliable semiempirical model was developed for phosphors: the 5d energy level for the $\mathrm{Ce^{3+}}$ and $\mathrm{Eu^{2+}}$ activator in various hosts was modeled. As clearly shown in Figure 3, the so-called total shift was evaluated as a function of several basic parameters such as crystal-field splitting energy, activator-anion ligand distance, coordination number, anion polarizability, site symmetry, and ionic size difference. The emission energy should be inferred from the total shift value if phosphors exhibited extremely dilute activator concentrations and Stokes shift was predictable. However, it is not possible to clearly define the emission peak wavelength, emission bandwidth, and quantum efficiency as a function of the characteristics described above for practical phosphors for use in LEDs. It is obvious that the information regarding the activator and anion ligand was insufficient to describe the phosphor performance. In this context, we introduced 32 descriptors, which covered not only the short-range information around the activator-anion polyhedra but also the information beyond. In order to deal with such a complicated, multiparameter problem along with a huge amount of existing data, statistical data mining would be of great assistance for the understanding and prediction task. Although a more detailed calculation strategy based on the DFT calculation might be more suitable for the understanding and prediction task, the cost would be prohibitive.



$$D(A) = \epsilon_c(A) + \frac{\epsilon_{\mathrm{cfs}}(A)}{r(A)} - 1890\ \mathrm{cm}^{-1}$$

$$\frac{\epsilon_c}{N} = \frac{\alpha_{\mathrm{sp}} e^2}{4\pi\epsilon_0}\left(\langle r^2\rangle_{5d} - \langle r^2\rangle_{4f}\right)\frac{1}{R_{\mathrm{eff}}^6},$$

$$\frac{1}{R_{\mathrm{eff}}^6} \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(R_i - \frac{1}{2}\Delta R)^6},$$

Figure 3. Simple schematic diagram and mathematical equations to evaluate the total shift (D) given by Dorenbos in ref 25. Such shift can be evaluated using several basic measurable parameters such as crystal-field splitting energy, activator-anion ligand distance, coordination number, anion polarizability, site symmetry, and ionic size difference.

The Dorenbos[25,26,34,35] model was succinctly represented in Figure 3, where, $r(A)$ expresses the ratio between crystal field splitting and crystal field shift, A stands for an arbitrary host, r represents the radial position of the electron in either the 5d or 4f orbital, and $\langle r^2 \rangle$ is the expectation value of r^2, α_{sp} is the polarizability, e is the elementary charge, and ε_0 is the permittivity of vacuum. The summation is over all N nearest coordinating anion ligands. R_i is the distance of the ligands from the activator ion, ΔR the effective Shannon ionic radii and ΔS is the Stokes shift.

The main emphasis in the above expressions was given to the activator−anion ligand distance, the activator−anion ligand coordination number and the anion polarizability, even though entire emission process is also dependent on the crystal structure, the site symmetry and the constituent element. The mathematical equations shown in Figure 3 worked best for a fixed activator site symmetry. The data set, which we collected from phosphors targeting the practical application to LEDs, showed a variety of host structures spanning from the space group $P2_1(4)$ to $Ia\overline{3}d$ (230). The structural equation model (SEM) should work for all different types of host structures.

According to the loading values for the factors and indicators appearing in the final SEM, which are listed in Supporting Information Table s2, the smaller A-A distance, smaller A-A coordination number, larger radius of C site element, and smaller electronegativity of the C-site element gave rise to an emission peak red-shift. To verify the findings shown in Supporting Information Table s2, two similar phosphors $Ba_2LiB_5O_{10}$:Eu^{2+} and BaB_8O_{13}:Eu^{2+},[36,37] from the data sets, were examined in terms of the local structure around the activator and emission peak wavelength. Figure 4 shows the A−X and A−A local structures along with the emission spectra for $Ba_2LiB_5O_{10}$:Eu^{2+} and BaB_8O_{13}:Eu^{2+}. The A−X local environments were similar, that is, the average distance was almost identical and the coordination number was exactly the same. The activator site symmetry for both phosphors was relatively low and also similar (1 for $Ba_2LiB_5O_{10}$:Eu^{2+} and 3 for BaB_8O_{13}:Eu^{2+}), so that the difference in site symmetry could not lead to a huge difference in emission energy based on the conventional theory. Despite the similarity in the A−X local structures, the emission spectra differed significantly, $Ba_2LiB_5O_{10}$:Eu^{2+} exhibited an emission band in the near UV range but BaB_8O_{13}:Eu^{2+} showed an orange-yellow color emission. This finding could not be explained definitively by conventional theory based on both the nephelauxetic effect and crystal field splitting. Assuming that the Stokes shift was similar, only a parameter in the Dorenbos model,[25,26] which can explain the huge difference in emission peak wavelength, could be the anion polarizability because the A−X local structure was identical. If the equation shown in Figure 3 had held and the Stokes shift had been similar, the change in the anion polarizability induced by the Li inclusion might have been a critical reason for the huge change in emission energy, because the other constituting elements were identical for both borate phosphors. It is certain that the small change in the anion polarizability induced by such a small number of Li incorporation was not only a reason for the huge change in emission wavelength, because the anion polarizability for LuF_3 and $LiLuF_4$ did not differ significantly from each other and the anion polarizability for YF_3 and $LiYF_4$ was also similar.[25]

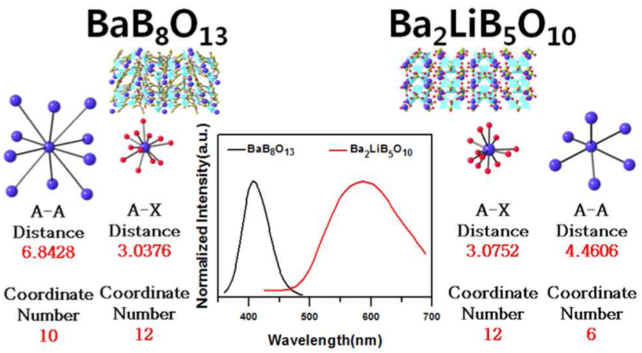


Figure 4. Emission spectra and A−X and A−A local structures for $Ba_2LiB_5O_{10}$:Eu^{2+} and BaB_8O_{13}:Eu^{2+}.

The CFA result based on the SEM proved to be powerful and economical in accounting for scientifically noninterpretable, complicated problems such as the $Ba_2LiB_5O_{10}$:Eu^{2+} and BaB_8O_{13}:Eu^{2+} cases. The emission peak energy difference could be explained by considering the differences in the A−A local environments between $Ba_2LiB_5O_{10}$:Eu^{2+} and BaB_8O_{13}:Eu^{2+}. The A−A distance and the coordination number for $Ba_2LiB_5O_{10}$:Eu^{2+} was much smaller than those for BaB_8O_{13}:Eu^{2+}. This finding was in perfect agreement with the regression weights (loadings) for the related factors and indicators in Supporting Information Table s2. Consequently, the smaller A−A distance and smaller A−A coordination number led to a red shift. The larger radius and smaller electronegativity of the C site element would lead to a red shift. In this regard, the red shift can also be well explained by Li incorporation at the C site. Thus, it is obvious that the structural equation model provided us with additional information on the top of what the theoretical approach would have given. Along with the change in the C site element by the Li inclusion, the A−A distance and the A−A coordination number also would influence the emission peak wavelength.

One might argue that the peak wavelength difference between both $Ba_2LiB_5O_{10}$:Eu^{2+} and BaB_8O_{13}:Eu^{2+} originated from the difference in Eu^{2+} concentration. But this would make no sense. In fact, the emission peak wavelength (PW) for both the phosphors was insensitive to the activator concentration and thereby never changed with the change in activator concentration.[36,37] It is thus, obvious that PWs obtained for both phosphors are not activator concentration-dependent but are a sort of intrinsic materials constant. Furthermore, it should be noted that the critical activator concentration (x_c) is not just an independent variable (an indicator) but is in fact a dependent variable (a target variable) just like a PW. Accordingly, it is baseless to argue that the difference in PW might originate from the difference in x_c.

5. CONCLUSION

To substantiate the theoretical approach, we employed a type of confirmatory factor analysis, which was proven to be powerful and allowed a better interpretation of the experimental data. In this regard, we collected materials information from the 75 Eu^{2+}-activated phosphors that have been reported thus far, and defined 32 descriptors, each of which described one of the material attributes. Using this 75 × 32 data set, we implemented a principal component analysis for data dimension reduction and the ensuing confirmatory factor analysis by setting up an appropriate structural equation model (SEM).

Using the reliable SEM, the underlying relationship between the emission peak wavelength (PW) of Eu^{2+}-activated phosphors and eight material descriptors was interpreted through four latent factors. The four latent factors were referred to as the A−X local environment factor (or small-scale environment factor), the A−A local environment factor (or large-scale environment factor), the anion trait factor, and the networking element trait factor. A causal effect from the A−A local environment factor to the A−X local environment factor was instrumental in the interpretation of phosphors that have never been definitively explained using conventional theory-based models.

The present work represents the successful application of a conventional strategy from the social sciences to solve a problem in materials science. With this process, we extracted valuable information from experimental data that we might have otherwise overlooked. When all available data are used, the understanding of complicated systems becomes much more efficient and economical compared with when we focus only on a theoretical approach.

■ ASSOCIATED CONTENT

Supporting Information

Additional tables that provide all data used in the manuscript, along with the reference list used for the data acquisition, brief description regarding the CFA and PLSR, and the model fit summary. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kssohn@sejong.ac.kr.

Author Contributions

W.B.P. and S.P.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013R1A2A2A04015089) and partly supported by partly by the Technology Innovation Program (or Industrial Strategic technology development program, 10050315, Discovery of narrow emission red phosphors for use in large color gamut and high power LEDs) funded By the Ministry of Trade, industry & Energy (MI, Korea).

■ ABBREVIATIONS

NSGA, nondominated sorting genetic algorithm; PSO, particle swarm optimization; CMS, combinatorial material search; HTE, high throughput experimentation; LED, light emitting diode; CFA, confirmatory factor analysis; SEM, structural equation model; PCA, principal component analysis; PLS, partial least-squares; ANN, artificial neural network; SVM, support vector machine; ICSD, inorganic compound structure database; SG, space group; SS, site symmetry; AM, activator site multiplicity; PW, peak wavelength; fwhm, full width at half-maximum; PCs, principal components

■ REFERENCES

(1) Engel, E.; Dreizler, R. M. *Density Functional Theory: An Advanced Course*; Springer: Heidelberg, 2011; pp 1−9.

(2) Yan, J.; Ning, L.; Huang, Y.; Liu, C.; Hou, D.; Zhang, B.; Huang, Y.; Tao, Y.; Liang, H. Luminescence and electronic properties of $Ba_2MgSi_2O_7:Eu^{2+}$: A combined experimental and hybrid density functional theory study. *J. Mater. Chem. C* **2014**, *2*, 8328−8332.

(3) Brito, H. F.; Felinto, M. C. F. C.; Hölsä, J.; Laamanen, T.; Lastusaari, M.; Malkamäki, M.; Novák, P.; Rodrigues, L. C. V.; Stefani, R. DFT and synchrotron radiation study of Eu^{2+}-doped $BaAl_2O_4$. *Opt. Mater. Express* **2012**, *2*, 420−431.

(4) Sohn, K.-S.; Cho, S. H.; Park, S. S.; Shin, N. Luminescence from two different crystallographic sites in $Sr_6BP_5O_{20}:Eu^{2+}$. *Appl. Phys. Lett.* **2006**, *89*, No. 051106.

(5) Fischer, C. C.; Tibbetts, K. J.; Morgan, D.; Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.* **2006**, *5*, 641−646.

(6) Curtarolo, S.; Morgan, D.; Persson, K.; Rodgers, J.; Ceder, G. Predicting crystal structures with data mining of quantum calculations. *Phys. Rev. Lett.* **2003**, *91*, No. 135503.

(7) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **2010**, *22*, 3762−3767.

(8) Park, W. B.; Singh, S. P.; Sohn, K.-S. Discovery of a phosphor for light emitting diode applications and its structural determination, $Ba(Si,Al)_5(O,N)_8:Eu^{2+}$. *J. Am. Chem. Soc.* **2014**, *136*, 2363−2373.

(9) Park, W. B.; Shin, N.; Hong, K.-P.; Pyo, M.; Sohn, K.-S. A new paradigm for materials discovery: Heuristics-assisted combinatorial chemistry involving parameterization of material novelty. *Adv. Funct. Mater.* **2012**, *22*, 2258−2266.

(10) Sohn, K.-S.; Lee, J. M.; Shin, N. A search for new red phosphors using a computational evolutionary optimization process. *Adv. Mater.* **2003**, *15*, 2081−2084.

(11) Allmann, R.; Hinek, R. The introduction of structure types into the inorganic crystal structure database ICSD. *Acta Crystallogr.* **2007**, *A63*, 412−417.

(12) https://icsd.fiz-karlsruhe.de/.

(13) Brown, T. A. *Confirmatory Factor Analysis for Applied Research*; The Guilford Press: New York, 2006; pp 1−11.

(14) Kline, R. B. *Principles and Practice of Structural Equation Modeling*, 3rd ed.; The Guilford Press: New York, 2010; pp 3−18.

(15) Amis, E. J. Combinatorial materials science: Reaching beyond discovery. *Nat. Mater.* **2004**, *3*, 83−85.

(16) Takeuchi, I.; Famodu, O. O.; Read, J. C.; Aronova, M. A.; Chang, K.-S.; Craciunescu, C.; Lofland, S. E.; Wuttig, M.; Wellstood, F. C.; Knauss, L.; Orozco, A. Identification of novel compositions of ferromagnetic shape-memory alloys using composition spreads. *Nat. Mater.* **2003**, *2*, 180−184.

(17) Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I.; Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **2006**, *5*, 909−913.

(18) Koinuma, H.; Takeuchi, I. Combinatorial solid-state chemistry of inorganic materials. *Nat. Mater.* **2004**, *3*, 429−438.

(19) Maier, W. F.; Stöwe, K.; Sieg, S. Combinatorial and high-throughput materials science. *Angew. Chem.* **2007**, *119*, 6122−6179; *Angew. Chem., Int. Ed.* **2007**, *46*, 6016−6067.

(20) Potyrailo, R.; Rajan, K.; Stoewe, K.; Takeuchi, I.; Chisholm, B.; Lam, H. Combinatorial and high-throughput screening of materials libraries: Review of state of the art. *ACS Comb. Sci.* **2011**, *13*, 579−633.

(21) Rajan, K. Materials informatics. *Mater. Today* **2005**, *8*, 38−45.

(22) Balachandran, P. V.; Broderick, S. R.; Rajan, K. Identifying the "inorganic gene" for high-temperature piezoelectric perovskites through statistical learning. *Proc. R. Soc. A* **2011**, *467*, 2271−2290.

(23) Rijsdijk, F. V.; Sham, P. C. Analytic approaches to twin data using structural equation models. *Brief. Bioinform.* **2002**, *3*, 119−133.

(24) Liu, B.; Fuente, A. d. l.; Hoeschele, I. Gene network inference via structural equation modeling in genetical genomics experiments. *Genetics* **2008**, *178*, 1763−1776.

(25) Dorenbos, P. 5d-level energies of Ce^{3+} and the crystalline environment. I. Fluoride compounds. *Phys. Rev. B* **2000**, *62*, 15640−15649.

(26) Dorenbos, P. 5d-level energies of Ce^{3+} and the crystalline environment. II. Chloride, bromide, and iodide compounds. *Phys. Rev. B* **2000**, *62*, 15650−15659.

(27) Pauling, L. *The Nature of the Chemical Bond and Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960; pp 65−105.

(28) Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chaleogenides. *Acta Crystallogr.* **1976**, *A32*, 751−767.

(29) Pitt, M. A.; Myung, I. J. When a good fit can be bad. *Trends Cogn. Sci.* **2002**, *6*, 421−425.

(30) Vásquez, S. O. Energy transfer processes in organized media. I. A crystal model for cubic sites. *J. Chem. Phys.* **1996**, *104*, 7652−7657.

(31) Vásquez, S. O. Energy transfer processes in organized media. II. Generalization of the crystal model for dipole−dipole interactions in cubic sites. *J. Chem. Phys.* **1997**, *106*, 8664−8671.

(32) Vásquez, S. O. Crystal model for energy-transfer processes in organized media: Higher-order electric multipolar interactions. *Phys. Rev. B* **1999**, *60*, 8575−8585.

(33) Vásquez, S. O. Energy-transfer processes in quasi-bidimensional crystal arrays. *Phys. Rev. B* **2001**, *64*, 125103.

(34) Dorenbos, P. 5d-level energies of Ce^{3+} and the crystalline environment. III. Oxides containing ionic complexes. *Phys. Rev. B* **2001**, *64*, No. 125117.

(35) Dorenbos, P. Relation between Eu^{2+} and Ce^{3+} f ↔ d-transition energies in inorganic compounds. *J. Phys.: Condens. Matter* **2003**, *15*, 4797−4807.

(36) Wang, Q.; Deng, D.; Xu, S.; Hua, Y.; Huang, L.; Wang, H.; Zhao, S.; Jia, G.; Li, C. Crystal structure and photoluminescence properties of Eu^{2+}-activated Ba_2LiB_5O10 phosphors. *Opt. Commun.* **2011**, *284*, 5315−5318.

(37) Sonekar, R. P.; Omanwar, S. K.; Moharil, S. V. Combustion synthesis and photoluminescence of Eu^{2+} doped BaB_8O_{13}. *Indian J. Pure Appl. Phys.* **2009**, *47*, 441−443.