

MedeA *HT-Descriptors*

High Throughput Flowchart Calculations

At-a-Glance

MedeA^{®1} *HT-Descriptors* is a tool within the *MedeA Environment* that can define, compute, exploit, and organize materials descriptors. For example, *HT-Descriptors* can identify and characterize layered compounds in a structural database, it can compute coordination geometries of selected atom types, and it provides tools to combine these experimental properties with computed data using mathematical expressions, thus generating sophisticated materials descriptors. These expressions are stored in catalogs which can be managed through the *HT-Descriptors* interface, ready for re-use on new structural datasets.

Key Benefits

- Efficient and practical approach to solve materials problems involving properties that cannot be computed directly
- Powerful tool to harvest descriptors from experimental crystal structure databases containing hundreds of thousands of compounds
- Leveraging the capability of *MedeA HT-Launchpad* to generate systematic and coherent sets of fundamental materials property data
- Ability to combine pre-calculated generic descriptors of entire structural databases (ICSD, Pearson, COD) with computed data
- Infrastructure to catalog and manage protocols for the calculation of descriptors

Many important materials properties are difficult to compute directly. Catalytic activity, stress-corrosion cracking, and lubrication are illustrative examples. Descriptors offer a path to link these phenomena to properties that can be obtained directly from a structural analysis, or which can be

¹ MedeA and Materials Design are registered trademarks of Materials Design, Inc.

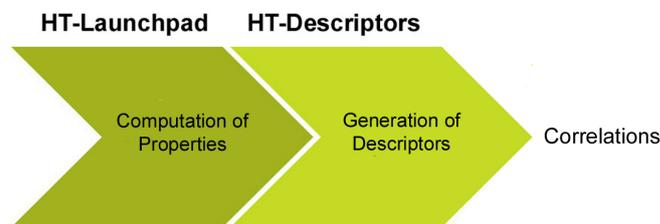


Figure 1: Role of MedeA HT-Descriptors as link between MedeA HT-Launchpad and the exploitation of correlations.

readily computed, such as the binding energy in compounds, the electronic density of states, elastic moduli, and the viscosity of fluids. A critical part of such an approach is the ability to combine a wide range of basic descriptors, such as the presence of layers, the width of channels, the coordination geometry of cations, or electronegativity differences (“ionicity”), with computed properties, such as the bulk modulus, vacancy formation energy, and the Debye temperature. *MedeA HT-Descriptors* facilitates the definition and exploitation of such descriptors in a convenient and sophisticated way.

‘Descriptors constructed from basic materials properties are like keys opening very heavy doors.’

An Illustrative Example

MedeA HT-Descriptors uses structure lists to associate descriptors with each structure. The following example illustrates this concept. The question is finding layered compounds which contain sulfur, and are likely to have a large band gap. To this end, two descriptors are defined, namely a topological descriptor identifying layered compounds containing sulfur and a second descriptor related to the ionicity of the compound. The latter is expressed as the difference between the highest and lowest electronegativity of all atoms in each compound. Using the interface of the *MedeA HT-Descriptors* module, the descriptors are defined

and stored in a catalog. Next, the descriptors are applied in a search of the ICSD and Pearson databases using *MedeA InfoMaticA*, resulting in 540 unique layered compounds containing sulfur. As expected, this list contains familiar compounds such as molybdenum sulfide. The results from this query are stored in a structure list for further investigation. Then, the electronegativity difference is computed for each of the 540 compounds, and they are sorted from highest to lowest value. In this particular example, the system with the highest ionicity is sodium dithionite, as illustrated in the adjacent figure.

ID	Name	Structural Formula	# atoms	Symmetry	Cell parameters	electronegativity range
87	ICSD.16642: Dithionium duo-disulfate(V)	Na2S2O8	16	P21c1	6.404 6.576 6.586 90 113.32 90	2.51
142	Pearson.1120279	NaHSO3NaHO	88	P21c1	5.847 21.214 7.516 90 103.82 90	2.51
154	Pearson.192697	Na2S2O6Na4HO	172	P21c1	11.429 11.279 17.37 90 94.794 90	2.51
155	ICSD.8098: Dithionium thiosulfate(V) pentahydrate	NaHSO3NaHO	88	P21c1	5.841 21.217 7.525 90 103.55 90	2.51
170	ICSD.24970: Sodium sulfate octahydrate	NaHSO4NaHO	62	P-1	7.889 6.096 6.2321 85.73 86.219 88.59	2.51
432	ICSD.78741: Dithium bis(hydrogensulfate) - sulfonic acid (1/1)	LiHSO4LiHSO3	84	Pccn	17.645 3.378 10.667 90 90 90	2.46
376	Pearson.105219	CaS2O4Na4O4	48	P-1	3.829 7.4391 11.2966 72.517 81.447 87.282	2.44
31	Pearson.30728	LiHSO4O2	10	P4/mmm	4.025 4.025 13.78 90 90 90	2.34
35	Pearson.309734	YbHSO4O2	10	P4/mmm	3.8 3.8 13.32 90 90 90	2.34
502	ICSD.48199: Lanthanum oxide trisulfate	LaHSO4O2	32	P6/mmm	5.841 5.851 16.003 90 90 90	2.34
159	Pearson.146234	LaHSO4	24	Cmca	13.215 3.945 3.938 90 90 90	2.34
28	Pearson.309737	CaHSO4O2	10	P4/mmm	4.02 4.02 13.62 90 90 90	2.32
522	Pearson.162201	CaHSO4Na4O4	184	Fddd	5.637 12.0469 26.7011 90 90 90	2.32
33	Pearson.309740	YbHSO4O2	10	P4/mmm	4 4 13.6 90 90 90	2.31
32	Pearson.309736	NaHSO4O2	10	P4/mmm	3.8 3.8 13.56 90 90 90	2.3
38	Pearson.307781	NaHSO4O2	10	P4/mmm	3.97 3.97 13.77 90 90 90	2.3
126	ICSD.248200: Cesium neodymium bis(sulfate)	NaHSO4O2Na12	40	P21c1	8.893 7.187 16.793 90 91.671 90	2.3
34	Pearson.309738	SmHSO4O2	10	P4/mmm	3.95 3.95 13.51 90 90 90	2.27
39	Pearson.304215	SmHSO4O2	10	P4/mmm	3.9 3.9 13.76 90 90 90	2.27
476	ICSD.80874: Cesium thiosulfate	CaS4O6Na4	16	Pmma	8.093 6.432 8.48 90 90 90	2.25

Figure 2: Structure list showing the electronegativity range as descriptor of layered compounds containing sulfur. In this example, Na-dithionite has the highest electronegativity range (ionicity). A subsequent calculation using MedeA VASP confirms that this compound is an insulator with a large band gap.

Characteristics

MedeA HT-Descriptors includes the following basic descriptors that can be used as building blocks to define new descriptors, using standard mathematical operators:

- Topological descriptors: cages, channels, layers including their dimensions, e.g. smallest channel diameter
- Coordination of atoms, including coordination number, type of coordination, e.g. octahedral, tetrahedral, square planar; deviation from ideal geometry; type of atoms in nearest neighbor coordination shell, distance to nearest neighbors, dispersion of distances; confidence in the coordination type defined by the separation between first and second nearest neighbor shells

- Choice between covalent and ionic radii for the determination of topological properties and coordination geometry
- Atomic properties, including atomic number, atomic mass, electronegativity, and valence
- Use of computed properties as arguments in defining descriptors
- Recursive definition: a new descriptor can contain other descriptors as arguments
- Tools for creating and managing catalogs of descriptors to be applied to new structure lists

MedeA HT-Descriptors applied to the structural databases of InfoMaticA relies on pre-computed properties, such as topological features (cages, channels, layers) since the computational analysis of hundreds of thousands of compounds contained in these databases requires significant computational effort. Thus, databases with these properties are delivered with the *MedeA* releases.

When operating on structure lists rather than full structural databases, *MedeA HT-Descriptors* computes these properties on the fly, thus giving the user great flexibility to include, for example, properties computed with *MedeA VASP* and *MedeA LAMMPS* in the expressions for descriptors. Structure lists are usually applied in the context of specific projects where the focus is on certain classes of compounds, rather than on all known structures stored in databases like ICSD and Pearson.

In summary, *MedeA HT-Descriptors* is a unique and extremely powerful tool for tackling complex materials problems by combining experimental structural data with computed properties obtained from the wealth of methods available in the *MedeA* environment.

Required Modules

- *MedeA Environment*
- *MedeA HT-Launchpad*

Related Modules

- *MedeA InfoMaticA*