機器學習於材料資訊的應用 Machine Learning on Material Informatics

陳南佑(NAN-YOW CHEN)

nanyow@narlabs.org.tw

楊安正(AN-CHENG YANG)

acyang@narlabs.org.tw

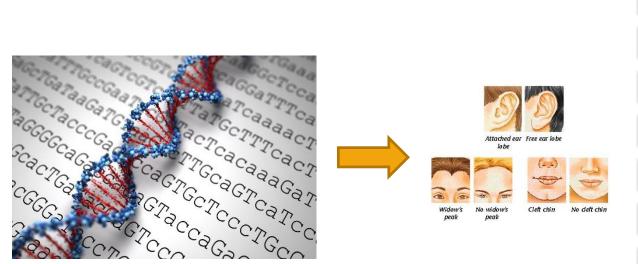
Data comes from Computational Materials Science

A new powerful approach to discover novel material

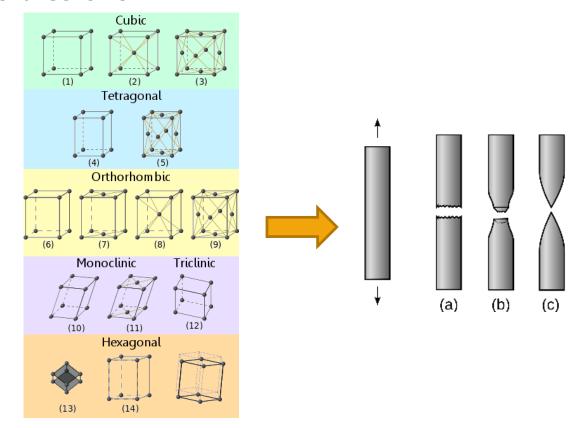


材料基因

Human Genome V.S. Material Genome

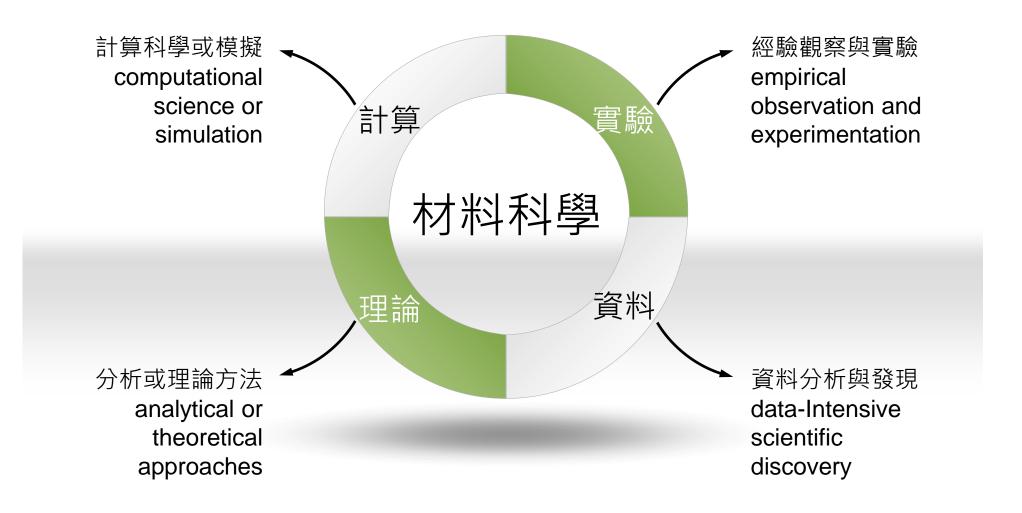


生物個體的外在表現叫作「表現型 (phenotype)」,由內在的「基因型 (genotype)」決定。改變表現型,首先要改變基因型。.

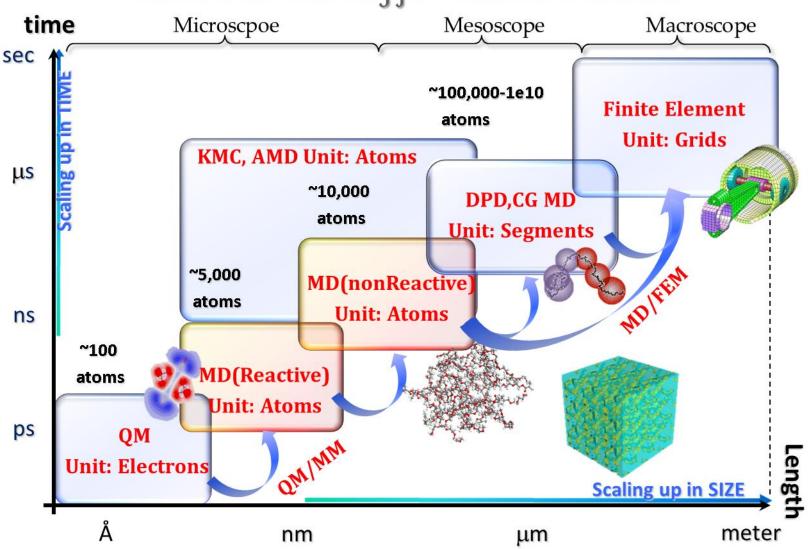


材料在巨觀尺度表現出來的材料特性是由微觀尺度下的原子組態所決定的。要改變材料特性,首先要改變微觀的原子組態。.

材料科學的分支



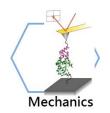
Multi-Scale Modeling for Materials Simulation



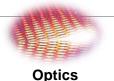




Properties you can get from calculation



Elastic constant Young's modus Stress-strain Hardness



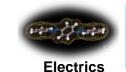
Birefringence Dielectric constant **Abbe Number** Reflective Index



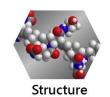
Glass transition temperature Coefficient of Thermal expansion

Thermal conductivity(Green Kubo)

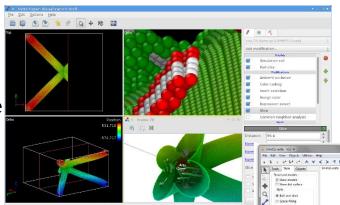
Phonon density **Solubility Parameters**



Band structure/gap Density of state



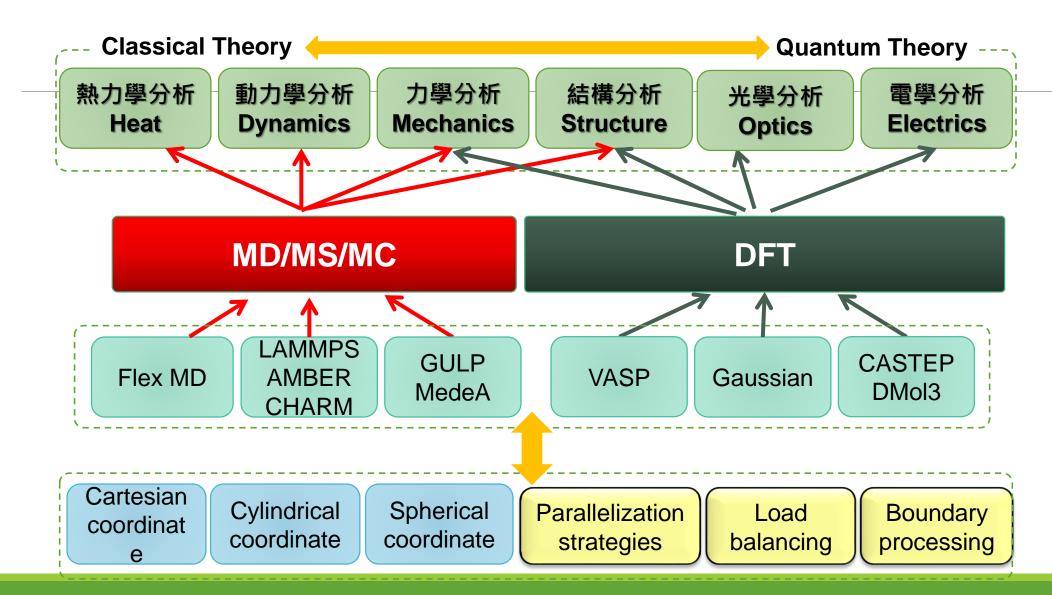
Radius distribution function Bond length, angle, dihedral angle **Gyration Radius** Atomic shear strain



Diffusion coefficient Viscosity (Green Kubo) Auto-correlation function



Molecular Modeling



Material Project



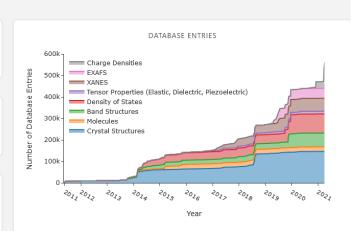


The Materials Project

Harnessing the power of supercomputing and state-of-the-art methods, the Materials Project provides open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials.

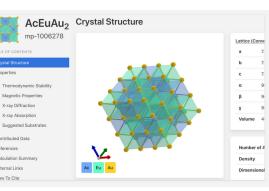
The Materials Project by the numbers

MATERIALS REGISTERED USERS 154,718 300,000+ INTERCALATION ELECTRODES CITATIONS 4,351 19,000 +MOLECULES CPU HOURS/YEAR 24,989 100 million



View detailed computed information about any material.

Every material in our database has a detailed page displaying all of its computed properties. By providing materials researchers with the information they need to design better, the Materials Project aims to accelerate innovation in materials research. See a random material.



Supercomputing

Supercomputing clusters at national laboratories provide the infrastructure that enables our computations, data, and algorithms to run at unparalleled speed. We principally use the Lawrence Berkeley National Laboratory's NERSC Scientific Computing Center and Computational Research Division, but we are also active with Oak Ridge's OLCF Argonne's ALCF and San Diego's SDSC



External Links How To Cite

About Material Project



software

By computing properties of all known materials, the Materials Project aims to guesswork from materials remove design in a variety of applications. Experimental research can be targeted to the most promising compounds from computational data sets. Researchers will be able to data-mine scientific trends in materials properties. By providing materials researchers the information they need to design better, the Materials Project aims to accelerate innovation in materials research.

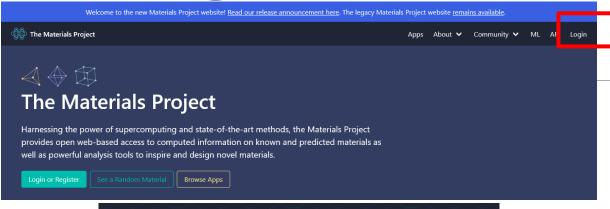
supercomputers

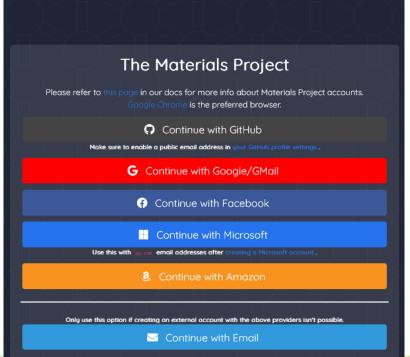
Supercomputing clusters at national laboratories provide the infrastructure that enables our computations, data, and algorithms to run at unparalleled speed. We principally use the Lawrence Berkeley National Laboratory's NERSC Scientific Computing Center and Computational Research Division, but we are also active with Oak Ridge's OLCF, Argonne's ALCF, and San Diego's SDSC.

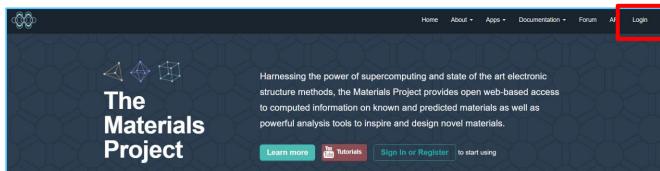
screening

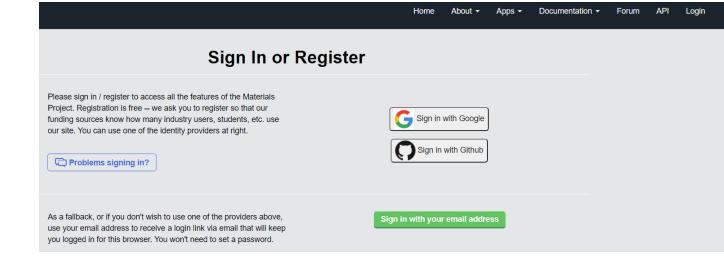
Computational materials science is now powerful enough that it can predict many properties of materials before those materials are ever synthesized in By scaling materials computations supercomputing over clusters, we have predicted several new battery materials which were made and tested in the lab. Recently, we have also identified new transparent conducting oxides and thermoelectric materials using this approach.

Login











也可以直接點

週期表元素





Explore Materials





Advanced Search Syntax





or property

He

× search

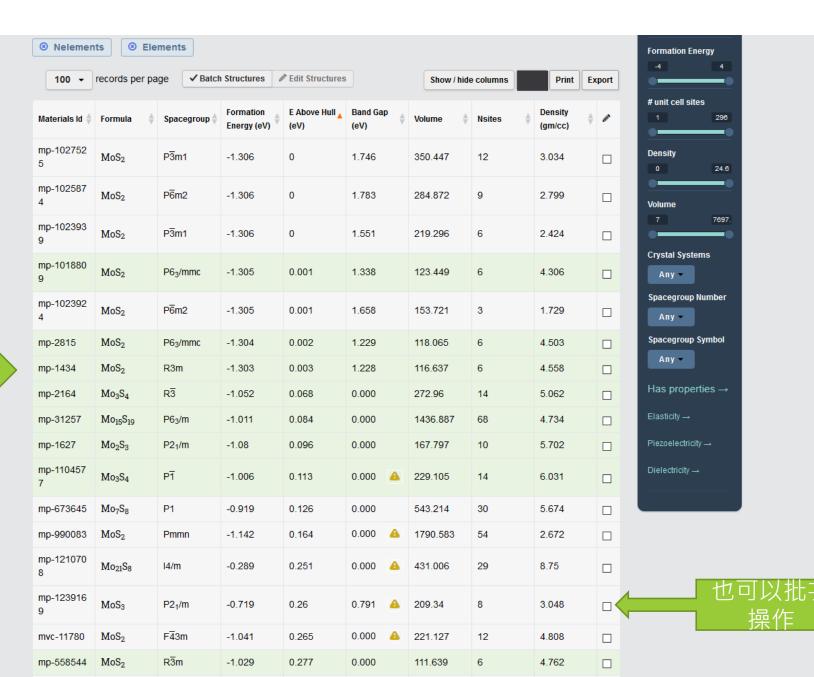


59 60 61 62 63 64 65 66 67 68 69 70 71 La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 Ac Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr

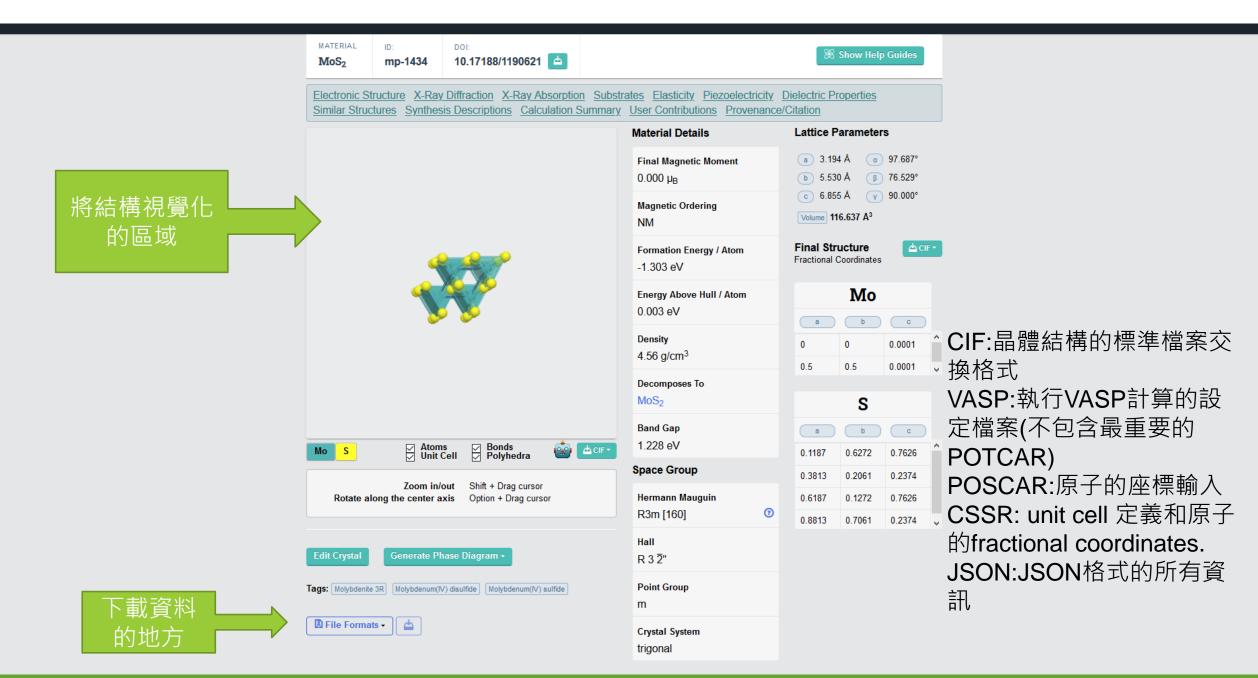
主要搜尋結果會呈現在這邊



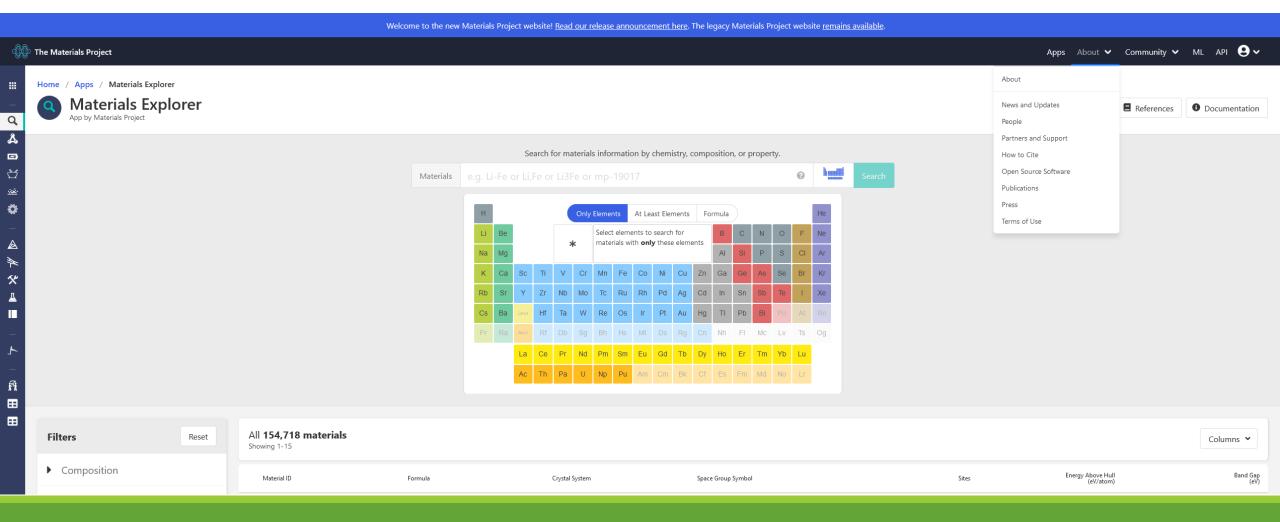




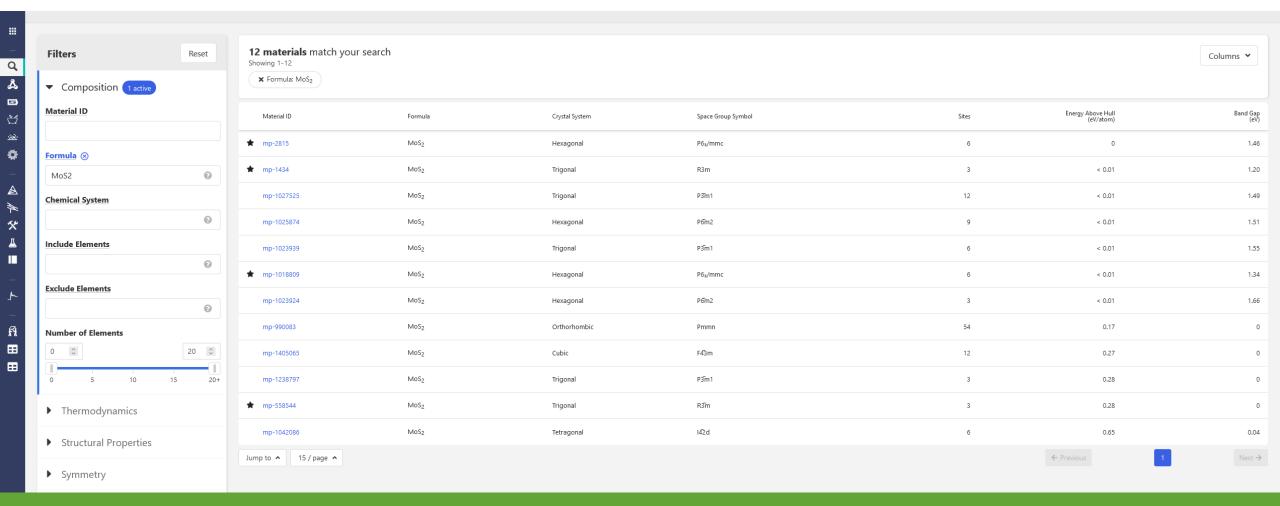
可以直接點進去看細節



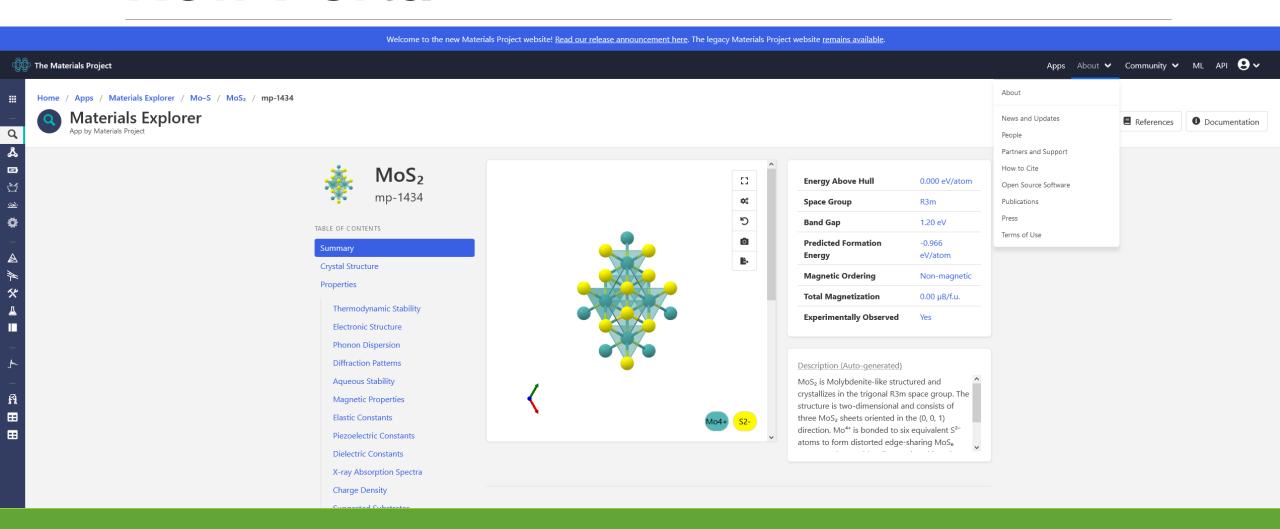
New Portal



New Portal

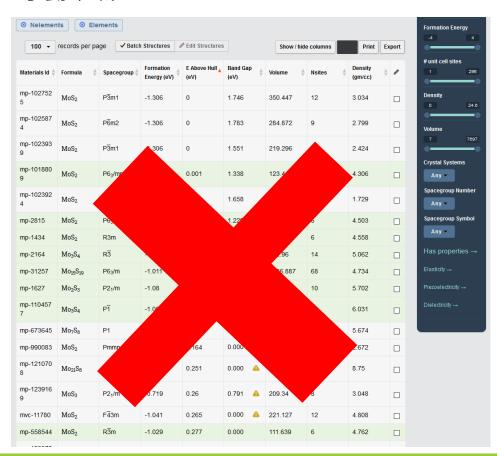


New Portal



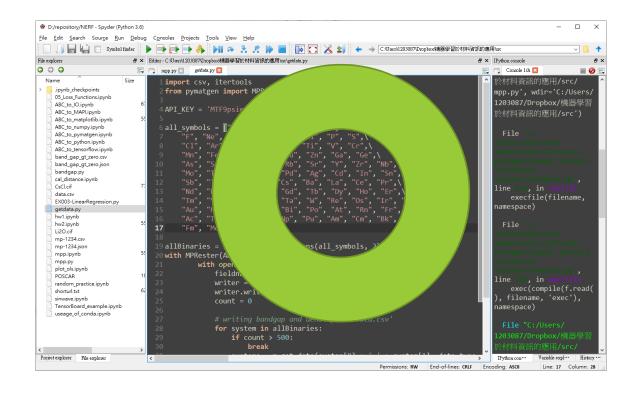
Get Data

手動下載



自動化下載

open Materials Application Programming Interface (API)



Open Source Software

- Pymatgen open source Python library for materials analysis
- Crystal Toolkit web app framework from the Materials Project which is now forming the basis of the Materials Project website
- ☐ FireWorks defining, managing, and executing scientific workflows
- Custodian job management framework that perform error checking, job management and error recovery
- Atomate provides "recipes" for performing complex materials science computations.
- Emmet is a toolkit of packages designed to build the Materials API.
- Maggma a framework to build complex data transformation pipelines from files to a REST API
- MPContribs provides API to contribute computational data to Materials Project.

Material Project

- □ Material Project提供了兩個工具供材料科學家使用他們的資料庫:
 - ▶ Materials Application Programming Interface (MAPI) · 資料釋出的協定。
 - ➤ Python Materials Genomics (pymatgen) materials analysis package,分析的工具包。
- □ MAPI透過REpresentational State Transfer (REST)架構,方便使用者透過http requests 存取資料。
- □ 在RESTful架構, 所有資訊都是整合至resources, 而resources也都透過 uniform resource identifier (URI)來指定網站上資料庫中的位置.
- □ Pymatgen (Python Materials Genomics)是一個針對材料科學的open-source Python library,特別是和Material Project高度整合。

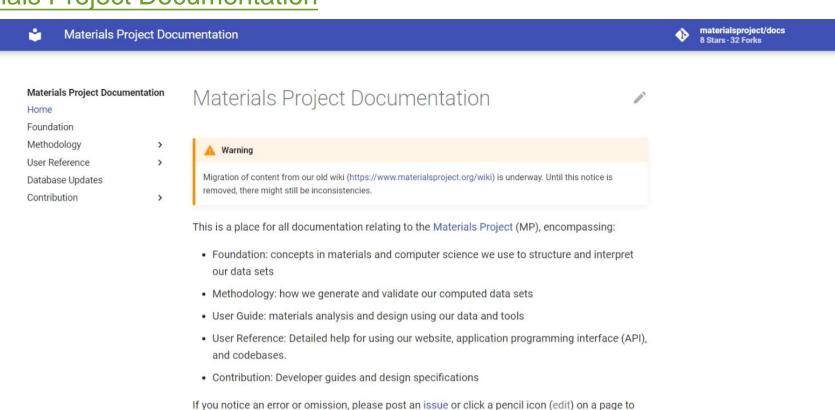
MAPI DOC

Materials Project Documentation

suggest an edit to us.

Materials Project Team

Thanks,



API keys

- 註冊登入後,可以到dashboard查看你的API KEY。
- □ 不要共用API KEY,如果有外流就重新產生一把。
- ☐ All MP https requests must supply API key as:
 - A x-api-key header, e.g., {'X-API-KEY': 'YOUR_API_KEY'} (recommended method)
 - As a GET (e.g., ?API_KEY=YOUR_API_KEY) or POST variable, e.g., {'API_KEY': 'YOUR_API_KEY'}
- □ 以GET方式,向Material Project網站要求material id 1234 的資源。

https://www.materialsproject.org/rest/v2/materials/mp-1234/vasp?API_KEY=YOUR_API_KEY

Find Your API-KEY



Apps About ♥ Community ♥

TABLE OF CONTENTS

The Materials Project API

Accessing Data

Documentation

Additional APIs

Getting Help

API Key



To make any request to the Materials Project API, you must use an API key. Your API key is generated for you automatically upon registering with the Materials Project website and is synced with the email you used to register.

Remember to keep your API key safe and to not share it with anyone you do not trust.

If you are logged in, you can always access your API key from this page or from your dashboard.

If you intend heavy API usage, you can give us a heads up by sending a message to heavy.api.use@materialsproject.org. This is not required, but may help us if we see unusual load on our servers. For large and/or long-running queries, we ask that you please make a local copy and only retrieve the data once.

Accessing Data

To use the API, you have three options:

1. You can use our first-party supported Python client. This is the recommend route. During the pre-release of the new website, this has to be installed manually via github.com/materialsproject/api.

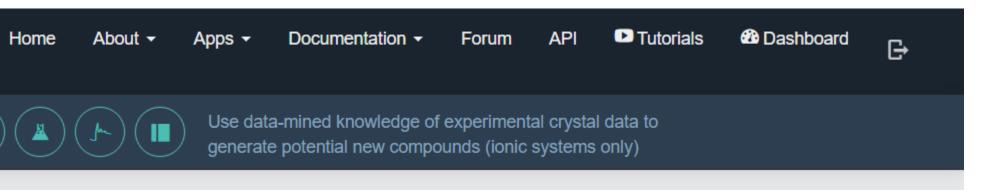
from mp_api.client import MPRester

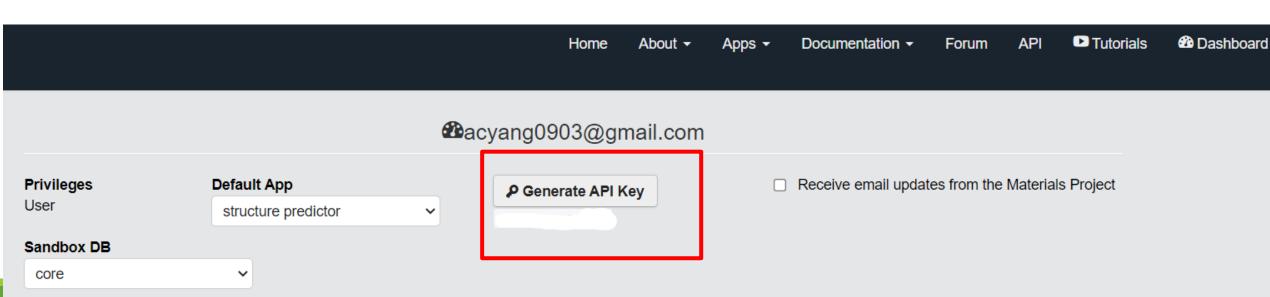
This may ultimately replace the legacy MPRester available in pymatgen .

- 2. You can demo the API interactively at api.materialsproject.org. Click the "Authorize" button, paste in your API key, and then click the appropriate section to try out a query.
- 3. Since this is a REST API, and offers a fully-compliant OpenAPI specification, it's possible to use the API with many libraries in many languages and environments, including JavaScript, MATLAB, Mathematica, etc. However, we do not offer first-party support for explaining how to do this, and you will have to follow the specification yourself.

Data Documents

Find Your API-KEY





Comparison of new API to legacy API

Comparison of new API to legacy API This table summarizes the differences between the new and legacy APIs for existing users. Legacy API New API Currently recommended for Early adopters Everyone else materialsproject.org/rest/v2 api.materialsproject.org Base URL Documentation api.materialsproject.org/docs mapidoc Specification OpenAPI-compliant specification available None available Our new API will be supported for the foreseeable Will be available for at least one year after new Support future once released API is finalized **Data Updates** Will receive new data updates included latest and Will be frozen at database release v2021.03.13 most accurate data Available at legacy.materialsproject.org/open **API Key** Available below

pip install pymatgen

No

from pymatgen.ext.matproj import MPRester

pip install mp-api (may be available in

from mp_api.client import MPRester

pymatgen at a later date)

Python client installation

Python client import code

contributed data

MPContribs integration for user

Uniform Resource Identifier (URI) design

Preamble URL

Request type Unique Identifier Data type

https://www.materialsproject.org/rest/v2/{request_type}[/{identifier}][/{parameters}]

- 1. preamble(前言、開場白)
 - 就是個網址而已,聲明這是一個https REST request, v2 表明這是version 2 of the MAPI。https://www.materialsproject.org/rest/v2/
- request_type

指名你request的是哪種操作或資訊,除了"materials"之外還有"tasks","battery", "reaction", "mpquery" and "api_check"。

identifier

根據request_type可以有不同的identifier,像是material id(e.g., mp-1234), formula(e.g. Fe2O3), chemical system("-" separated list of elemments, e.g., Li-Fe-O), or task id。

4. parameters

根據request_type和identifier決定的參數,視情況可能會有多個parameters。

像是vasp表示的是computational data,exp表示的是experimental data。

而vasp/energy表示的是computational data中的energy性質。

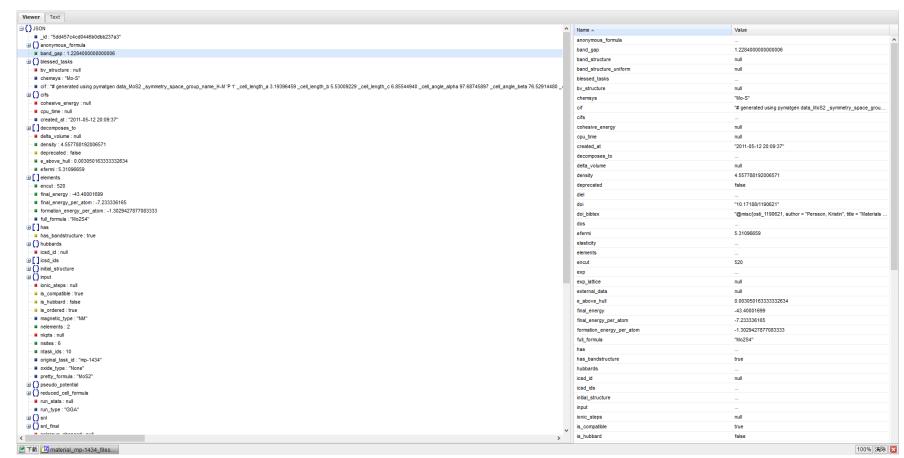
Sample JSON output

- □ 現階段回傳的物件都是以JavaScript Object Notation (JSON)回傳, 目前還未支援XML。
- □ 直接在網址列輸入http request,會接在網頁呈現。
- □ 可以另存網頁得到JSON檔案。

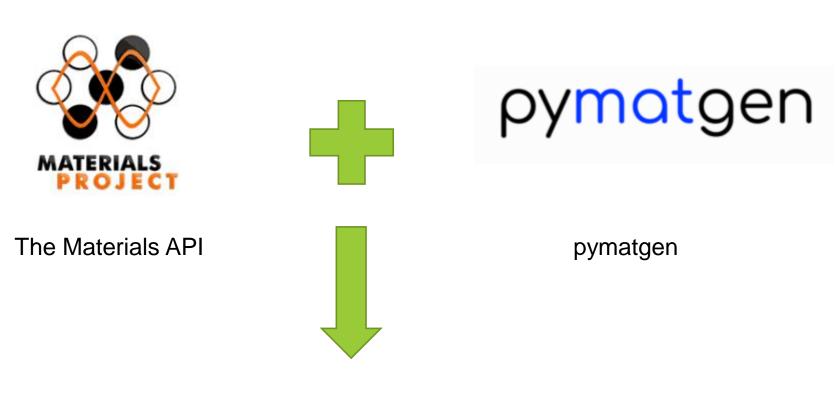
{"response": [{"energy": -26.91801468, "energy_per_atom": -4.48633578, "volume": 116.92379066659765, "formation_energy_per_atom": -0.48230172, "nsites": 6, "unit_cell_formula": {"Lu": 2.0, "Al": 4.0}, "pretty_formula": "LuAl2", "is_hubbard": false, "elements": ["Lu", "Al"], "nelements": 2, "e_above_hull": 0, "hubbards": {}, "is_compatible": true, "spacegroup": {"symprec": 0.1, "source": "spglib", "symbol": "Fd-3m", "number": 227, "point_group": "m-3m", "crystal_system": "cubic", "hall": "F 4d 2 3 -1d"}, "task_ids": ["mp-940234", "mp-925833", "mp-940654", "mp-1234", "mp-1258536", "mp-1438947", "mp-1678516", "mp-1803625", "mp-1593631"], "band gap": 0.0, "density": 6.502480539556319, "icsd id": null, "icsd ids": [608375, 57958, 608376, 608372, 608371, 608370], "cif": "# generated using pymatgen\ndata LuAl2\n symmetry space group name H-M 'P 1'\n cell length a 5.48873961\n cell length b 5.48873961\n cell length c 5.48873961\n cell angle alpha 60.00000000\n cell angle beta 60.00000000\n cell angle gamma 60.00000000\n symmetry Int Tables number 1\n chemical formula structural LuAl2\n chemical formula sum 'Lu2 Al4'\n cell volume 116.92379067\n cell formula units Z 2\nloop \n symmetry equiv pos site id\n symmetry equiv pos as xyz\n 1 'x, y, z'\nloop \n atom site type symbol\n atom site label\n atom site symmetry multiplicity\n atom site fract x\n _atom_site_fract_y\n_atom_site_fract_z\n_atom_site_occupancy\n_Lu_Lu0_1 0.25000000 0.25000000 0.25000000 1\n_Lu_Lu1_1 0.00000000 0.000000000 0.000000000 1\n_Al_Al2_1 0.62500000 0.12500000 0.62500000 1\n Al Al3 1 0.62500000 0.62500000 0.12500000 1\n Al Al4 1 0.12500000 0.62500000 0.62500000 1\n Al Al5 1 0.62500000 0.62500000 0.62500000 1\n", "total_magnetization": 0.00022485 "material id": "mp-1234", "oxide type": "None", "tags": ["Aluminium lutetium (2/1)"], "elasticity": {"G Reuss": 61.0, "G VRH": 61.0, "G Voigt": 62.0, "G Voigt Reuss Hill": 61.0, "K Reuss": 82.0, "K VRH": 82.0, "K Voigt": 82.0, "K Voigt Reuss Hill": 82.0, "elastic anisotropy": 0.04, "elastic tensor": [[174.0, 36.0, 36.0, 0.0, 0.0], [36.0, 174.0, 36.0, 0.0, 0.0], [36.0, 174.0, 36.0, 0.0], [36.0, 36.0, 174.0, 0.0, 0.0], [0.0, 0.0] 0.0, 57.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0, 57.0, 0.0], [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 57.0]], "homogeneous_poisson": 0.2, "poisson_ratio": 0.2, "universal_anisotropy": 0.04, "elastic_tensor_original": $\lceil [173.8039399290068, 36.29190678313171, 36.29190678313171, 0.0, 0.0, 0.0 \rceil, [36.29190678313171, 173.8039399290068, 36.29190678313171, 0.0, 0.0, 0.0 \rceil, [36.29190678313171, 0.0, 0.0, 0.0], [36.29190678313171, 0.0, 0.0, 0.0], [36.29190678313171, 0.0, 0.0, 0.0], [36.29190678313171, 0.0, 0.0, 0.0], [36.29190678313171, 0.0, 0.0, 0.0], [36.29190678313171, 0.0, 0.0, 0.0], [36.29190678313171, 0.0, 0.0, 0.0], [36.29190678313171, 0.0, 0.0], [36.29190678313171, 0.0, 0.0], [36.29190678313171, 0.0, 0.0], [36.29190678313171, 0.0, 0.0], [36.29190678313171, 0.0, 0.0], [36.29190678313171, 0.0], [36$ 0.0, 0.0], [0.0, 0.0, 0.0, 56.801403605413576, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0, 0.0, 56.801403605413576, 0.0], [0.0, 0.0, 0.0, 0.0, 0.0, 56.801403605413576]], "compliance tensor": [[6.2, -1.1, -1.1, 0.0, -0.0, 0.0], [-1.1, 6.2, -1.1, -0.0, 0.0, -0.0], [-1.1, -1.1, 6.2, -0.0, -0.0, 0.0], [0.0, -0.0, -0.0, 17.6, 0.0, -0.0], [-0.0, 0.0, -0.0, 0.0, -0.0], [0.0, -0.0, -0.0, -0.0, -0.0], [0.0, -0.0, -0.0], [0.0, -0.0, -0.0], [0.0, -0.0, -0.0], [0.0, -0.0, -0.0], [0.0, -0.0], [0.0, -0.0, -0.0], [0.0, -0.0, -0.0], [0.0, -0.0, -0.0], [0 earth element"], "nsites": 6}, "piezo": null, "diel": null, "deprecated": false, "full_formula": "Lu2Al4"}], "valid_response": true, "created_at": "2021-05-31T20:46:37.838142", "version": {"db": "2021_05_13", "pymatgen": "2022.0.8", "rest": "2.0"}, "copyright": "Materials Project, 2021"}

Online jsonviewer http://jsonviewer.stack.hu/





Integration with pymatgen

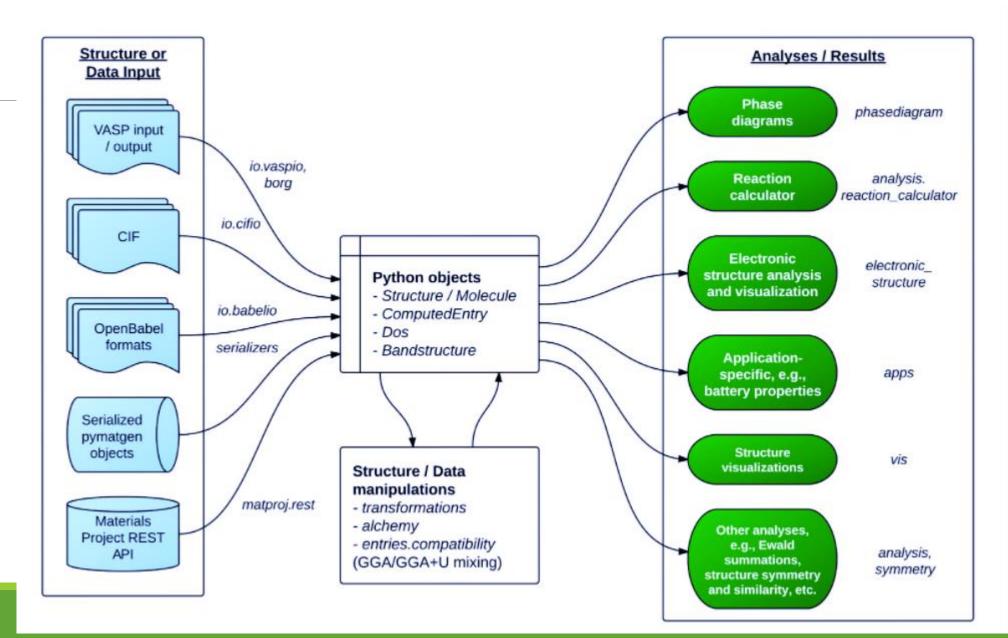


Powerful Materials
Analytics Tool

Introduction to Pymatgen

- □ 提供Element, Site, Molecule, Structure objects多種class,方便用戶進一步操作。
- □ 支援VASP, ABINIT, CIF, Gaussian, XYZ等多種檔案格式
- □ 提供多種分析模組phase diagrams, Pourbaix diagrams, diffusion analyses, reactions, Electronic structure, density of states和band structure。
- □ 和Materials Project REST API, Crystallography Open Database and other external data sources的API整合。
- □ As of 2023, pymatgen only supports Python 3.7 and above latest version is v2023.5.31

typical workflow for pymatgen

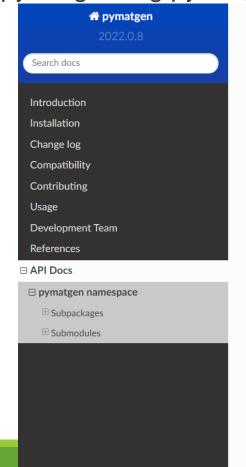


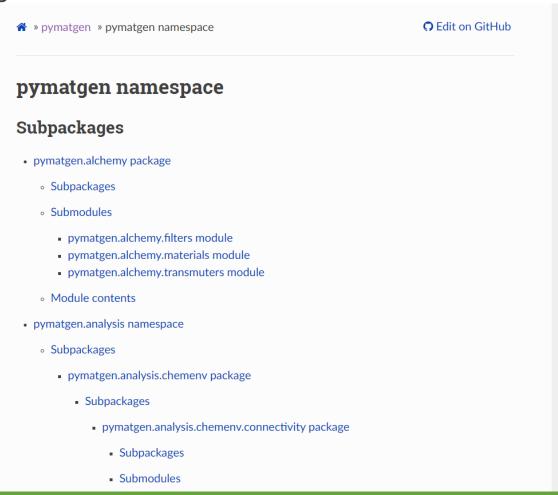
Compatibility of pymatgen v2022.0.0

```
from pymatgen import Composition # now "from pymatgen.core.composition import Composition"
from pymatgen import Lattice # now "from pymatgen.core.lattice import Lattice"
from pymatgen import SymmOp # now "from pymatgen.core.operations import SymmOp"
from pymatgen import DummySpecie, DummySpecies, Element, Specie, Species # now "from
pymatgen.core.periodic_table ..."
from pymatgen import PeriodicSite, Site # now "from pymatgen.core.sites ..."
from pymatgen import IMolecule, IStructure, Molecule, Structure # now "from
pymatgen.core.structure ..."
from pymatgen import ArrayWithUnit, FloatWithUnit, Unit # now "from pymatgen.core.units ..."
from pymatgen import Orbital, Spin #now "from pymatgen.electronic_structure.core ..."
from pymatgen import MPRester # now "from pymatgen.ext.matproj ..."
```

Pymatgen API DOC

https://pymatgen.org/pymatgen.html





pymatgen namespace

- pymatgen.alchemy
- pymatgen.analysis
- pymatgen.apps
- pymatgen.cli
- pymatgen.command_line
- pymatgen.core
- pymatgen.electronic_structure
- pymatgen.entries
- pymatgen.ext

pymatgen namespace

- pymatgen.io
- pymatgen.optimization
- pymatgen.phonon
- pymatgen.symmetry
- pymatgen.transformations
- pymatgen.util
- pymatgen.vis

pymatgen.core

- pymatgen.core.bonds module
- pymatgen.core.composition module
- pymatgen.core.ion module
- pymatgen.core.lattice module
- pymatgen.core.libxcfunc module
- pymatgen.core.molecular_orbitals module
- pymatgen.core.operations module
- pymatgen.core.periodic_table module

pymatgen.core

- pymatgen.core.sites module
- pymatgen.core.spectrum module
- pymatgen.core.structure module
- pymatgen.core.surface module
- pymatgen.core.tensors module
- pymatgen.core.trajectory module
- pymatgen.core.units module
- pymatgen.core.xcfunc module

pymatgen.ext

- pymatgen.ext.cod module
- pymatgen.ext.crystalsai module
- pymatgen.ext.jhu module
- pymatgen.ext.matproj module
- pymatgen.ext.optimade module

pymatgen.electronic_structure

- pymatgen.electronic_structure.bandstructure module
- pymatgen.electronic_structure.boltztrap module
- pymatgen.electronic_structure.boltztrap2 module
- pymatgen.electronic_structure.cohp module
- pymatgen.electronic_structure.core module
- pymatgen.electronic_structure.dos module
- pymatgen.electronic_structure.plotter module

- pymatgen.io.abinit
- pymatgen.io.cp2k
- pymatgen.io.exciting
- pymatgen.io.feff
- pymatgen.io.lammps
- pymatgen.io.lobster
- pymatgen.io.qchem
- pymatgen.io.vasp

- pymatgen.io.adf module
- pymatgen.io.aiida module
- pymatgen.io.ase module
- pymatgen.io.atat module
- pymatgen.io.babel module
- pymatgen.io.cif module
- pymatgen.io.cssr module
- pymatgen.io.cube module

- pymatgen.io.fiesta module
- pymatgen.io.gaussian module
- pymatgen.io.jarvis module
- pymatgen.io.lmto module
- pymatgen.io.nwchem module
- pymatgen.io.phonopy module
- pymatgen.io.prismatic module

- pymatgen.io.pwscf module
- pymatgen.io.shengbte module
- pymatgen.io.wannier90 module
- pymatgen.io.xcrysden module
- pymatgen.io.xr module
- pymatgen.io.xyz module

- pymatgen.analysis.adsorption module
- pymatgen.analysis.bond_dissociation module
- pymatgen.analysis.bond_valence module
- pymatgen.analysis.cost module
- pymatgen.analysis.diffusion_analyzer module
- pymatgen.analysis.dimensionality module
- pymatgen.analysis.energy_models module
- pymatgen.analysis.eos module
- pymatgen.analysis.ewald module

- pymatgen.analysis.excitation module
- pymatgen.analysis.find_dimension module
- pymatgen.analysis.fragmenter module
- pymatgen.analysis.functional_groups module
- pymatgen.analysis.graphs module
- pymatgen.analysis.hhi module
- pymatgen.analysis.interface module
- pymatgen.analysis.interface_reactions module
- pymatgen.analysis.local_env module
- pymatgen.analysis.molecule_matcher module

- pymatgen.analysis.molecule_structure_comparator module
- pymatgen.analysis.nmr module
- pymatgen.analysis.path_finder module
- pymatgen.analysis.phase_diagram module
- pymatgen.analysis.piezo module
- pymatgen.analysis.piezo_sensitivity module
- pymatgen.analysis.pourbaix_diagram module
- pymatgen.analysis.prototypes module
- pymatgen.analysis.quasiharmonic module

- pymatgen.analysis.reaction_calculator module
- pymatgen.analysis.structure_analyzer module
- pymatgen.analysis.structure_matcher module
- pymatgen.analysis.substrate_analyzer module
- pymatgen.analysis.surface_analysis module
- pymatgen.analysis.thermochemistry module
- pymatgen.analysis.transition_state module
- pymatgen.analysis.wulff module
- pymatgen.analysis.xps module