

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

The code corresponding to our implementation is available in the Github repository <https://github.com/AleFalla/Properties-to-molecules-Inverse-Mapping>. The main packages used to run the python scripts were python (3.7.4), numpy (1.21.4), scipy (1.7.3), scikit-learn (1.0.1), pandas (1.3.4), pytorch (1.12.1), pytorch-lightning (1.5.10), matplotlib (3.5.0), and rmsd (1.4). The codes used to optimize the 3D molecular structures and compute the quantum-mechanical properties were Open Babel (v3.1.1), DFTB+ (v23.1), and FHI-aims (v221103).

Data analysis

The main packages used to run the python scripts were python (3.7.4), numpy (1.21.4), scipy (1.7.3), scikit-learn (1.0.1), pandas (1.3.4), pytorch (1.12.1), pytorch-lightning (1.5.10), matplotlib (3.5.0), and rmsd (1.4).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The QM7-X data that support the findings of this study are available in ZENODO with the identifier doi:10.5281/zenodo.4288677. The preprocessed data to make

our implementation work are provided in the Github repository <https://github.com/AleFalla/Properties-to-molecules-Inverse-Mapping>. The authors declare that all data supporting the findings of this study are available within the paper and its supplementary information files.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We have used 40,988 equilibrium conformations of molecules with up to seven heavy atoms contained in QM7-X dataset. This set of molecules was systematically constructed to span the region in chemical space corresponding to small molecules.
Data exclusions	We have considered only equilibrium molecules with up to seven heavy atoms including C, N, and O.
Replication	In the GitHub repository ( <a href="https://github.com/AleFalla/Properties-to-molecules-Inverse-Mapping">https://github.com/AleFalla/Properties-to-molecules-Inverse-Mapping</a> ), we are providing the preprocessed data to reproduce our findings.
Randomization	The data was randomly selected to train the machine learning models. This ensures the development of an unbiased model.
Blinding	N/A

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

# Plants

---

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A