CSP Core Dictionary

Introduction

Summary

This is the CSP dictionary for describing predicted crystal structures and the methods, parameters and workflows used to calculate these.

Table of contents

1. Input Chemical System

This section specifies the atoms used in inorganic CSP or the input molecules for molecular crystal generation.

Category	Data Field	Type	Definition	Constraints	Units	Example
Input	`name`	char	See name_common and name_systematic from Core CIF dictionary.	Free Text		urea hydrate
Input	`composition_calculation`	char	"fixed" or "variable" composition calculation	- Fixed - Variable		Fixed
Input	`composition_coefficients`	list	List of possible compositions for fixed-composition calculations or extremes for variable-composition simulations	`List[PositiveInt]`		`[1 1]` `[2 1]`
Input	`maximum_number_of_components`	numb	The maximum number of components (atoms or molecules) in the unit cell	1:		4
Input	`minimum_number_of_components`	numb	The minimum number of components (atoms or molecules) in the unit cell	1:		2
Input Atoms	`types`	list	List of atomic species defining the composition	`List[str]` or `List[PositiveInt]`		`[Mg O]` (atomic symbols), `[12 8]` (Atomic number)
Input Molecule	`number`	char	Molecule component index.	1:		1
Input Molecule	`identifier`	char	Label used to identify the molecule.	Free Text		urea, water
Input Molecule	`smiles`	char	SMILES of the component.	-		C(=O)(N)N
Input Molecule	`molecule_number`	char	Molecule component index for each atom.	1:		1
Input Molecule	`molecule_identifier`	char	Label used to identify the molecule for each atom.	Free Text		urea, water
Input Molecule	`atom_label`	char	Label of atom in the component.	Free Text		C1

Additional details on atoms in molecule ad their connectivity can be specified through the CIF Chemical dictionary, available at:
https://www.iucr.org/__data/iucr/cifdic_html/1/cif_core.dic/index.html

Examples

Inorganic CSP input with fixed stoichiometry:

```
_csp.input.name Ferrosilite
_csp.input_atoms.types Fe Si O
_csp.input.composition_calculation fixed
_csp.input.composition_coefficients [1 1 3]
```

Inorganic CSP input with variable stoichiometry:

```
_csp.input.name Hypersthene
_csp.input_atoms.types Fe Mg Si O
_csp.input.composition_calculation variable
_csp.input.composition_coefficients [[1 0 1 3] [0 1 1 3]]
_csp.input.minimum_number_of_components 2
_csp.input.maximum_number_of_components 10
```

This implies that resulting structures will have formula *x*(FeSiO3)+*y*(MgSiO3) with *2<x+y<10*.

Multi-component molecular crystal CSP with fixed stoichiometry:

```
_csp.input.name Urea_Hydrate

# Molecules
loop_
    _csp.input_molecule.number
    _csp.input_molecule.identifier
    _csp.input_molecule.smiles
    _chemical.name_common
    1 WAT O      water
    2 URE OCN(N) urea

# Atoms in molecules
loop_
    _csp.input_molecule.molecule_number
    _csp.input_molecule.molecule_identifier
    _csp.input_molecule.atom_label
    _chemical.conn_atom.number
    _chemical.conn_atom.type_symbol
    _chemical.conn_atom.charge
    1 WAT O1 1  O -0.800000
    1 WAT H1 2  H  0.400000
    1 WAT H1 3  H  0.400000
    2 URE O1 4  O -0.613359
    2 URE C1 5  C  0.880229
    2 URE N1 6  N -0.923545
    2 URE N2 7  N -0.923545
    2 URE H1 8  H  0.395055
    2 URE H2 9  H  0.395055
    2 URE H3 10 H  0.395055
    2 URE H4 11 H  0.395055

# Bonds
loop_
    _chemical.conn_bond.atom_1
    _chemical.conn_bond.atom_2
    _chemical.conn_bond.type
    1 2  sing
    1 3  sing
    4 5  doub
    5 6  sing
    5 7  sing
    6 8  sing
    6 9  sing
    7 10 sing
    8 11 sing

_csp.input.composition_calculation "fixed"
_csp.input.composition_coefficients  [2 1] # Indexes from molecule section (2 water molecules and one urea)
```

`composition_coefficients` here refers to the molecule number. Worthy of note the use of the `Chemical` dictionary in defining the molecules.

Variable stoichiometry search can be specified in the same way as for inorganic systems:

```
...
_csp.input.composition_calculation "variable"
_csp.input.composition_coefficients  [[1 0] [0 1]]
_csp.input.maximum_number_of_components 4
_csp.input.minimum_number_of_components 2
```

For metal-organic systems, the `input_molecule` and `Chemical` dictionaries can be used specifying metallic atoms as "individual molecules":

```
_csp.input.name "(mi-tricyanomethanide)-silver"

# Molecules
loop_
    _csp.input_molecule.number
    _csp.input_molecule.identifier
    _chemical.name_common
    1 Metal Silver
    2 c4n3  tricyanomethanide

# Atoms in molecules
loop_
    _csp.input_molecule.molecule_number
    _csp.input_molecule.molecule_identifier
    _csp.input_molecule.atom_label
    _chemical.conn_atom.number
    _chemical.conn_atom.type_symbol
    1 Metal Ag1 1  Ag
    2 c4n3  C1  2  C
    2 c4n3  C2  3  C
    2 c4n3  C3  4  C
    2 c4n3  C4  5  C
    2 c4n3  N1  6  N
    2 c4n3  N2  7  N
    2 c4n3  N3  8  N

# Bonds
loop_
    _chemical.conn_bond.atom_1
    _chemical.conn_bond.atom_2
    _chemical.conn_bond.type
    1 6  sing
    1 7  sing
    1 8  sing
    2 3  doub
    2 4  sing
    2 5  sing
    3 6  doub
    4 7  trip
    5 8  trip


_csp.input.composition_calculation "fixed"
_csp.input.composition_coefficients  [1 1]
```

2. Structure Generation Methods

This section helps delineate the space search range and specify the parameters used for different methods.

2.1 General Fields

Category `_csp.structure_generation_[]` : Category for structure generation methods.

Category	Data Field	Type	Definition	Constraints	Units	Example
Structure Generation	`space_group_number_list`	char/numb/list	Space group selection could be “all” or a subset (list) specifying which spacegroups were used.	Either "all" or list of sg numbers		`[14 2 15 61 19 4 33 29 5 1]`
Structure Generation	`method`	char/list	Structure generation method or list of methods.	- Evolutionary Algorithm (Sec. 2.2) - Particle		Simulated Annealing

Category	Data Field	Type	Definition	Constraints	Units	Example
				Swarm Optimisation (Sec. 2.3) - Simulated Annealing (Sec. 2.4) - Monte Carlo Parallel tempering (Sec. 2.5) - Random Sampling (Sec. 2.6) - Analogue Templates - Other		
Structure Generation	`software`	char	Name of the software used for structure generation.	Free text		
Structure Generation	`software_citation`	char	Details of the software used for structure generation. Either URL to webpage or DOI of the related publication.	Free text		
Structure Generation	`software_version`	char	Version of software used for structure generation.	Free text		
Structure Generation	`density_lower_limit`	numb	Minimum Cell Density.	>0	kg m^{-3}	800
Structure Generation	`density_upper_limit`	numb	Maximum Cell Density.	>0	kg m^{-3}	1400
Structure Generation	`reference_temperature`	numb	Reference temperature for finite-temperature simulations.	>0	K	0
Structure Generation	`reference_pressure`	numb	Reference pressure for variable cell calculations.	>0	Pa	100000
Structure Generation	`stopping_criteria`	char/list	List of rules for stopping the generation of new structures.	Free text		"Max Structures", "Low-Energy Structures Unchanged"
Structure Generation	`stopping_criteria_max_structures_evaluated`	numb	The maximum total number of unique crystal structures that will be generated and evaluated	>0		10000

Category	Data Field	Type	Definition	Constraints	Units	Example
			during the search.			
Structure Generation	`stopping_criteria_iterations_without_improvement`	numb	The maximum number of consecutive iterations (generations, MC steps, etc.) where the global minimum (or the lowest few structures) does not change.	>0		50
Structure Generation	`stopping_criteria_energy_range`	numb	An energy threshold for the selection of low-energy structures to be considered in the convergence criteria.	>0	kJ mol^{-1}	5
Structure Generation	`stopping_criteria_structures_range`	numb	The number of low-energy structures to be considered in the convergence criteria.	>0		1000

2.2 Evolutionary Algorithms

Category `_csp.evolutionary_algorithm_[]` : Subgroup for CSP Structure Generation methods that use Evolutionary Algorithms. For these fields to be used, the `_csp.structure_generation.method` must include "Evolutionary Algorithm".

Category	Data Field	Type	Definition	Constraints	Units	Example
Evolutionary Algorithms	`population_size`	numb	The number of candidate structures in each generation.	>0		100
Evolutionary Algorithms	`initial_population_size`	numb	The number of candidate structures in the first generation.	>0		50
Evolutionary Algorithms	`number_of_generations`	numb	The maximum number of evolutionary cycles the algorithm will run before termination (unless other stopping criteria are met).	>0		50
Evolutionary Algorithms	`nextgen_structure_selection`	numb	The number of individuals that survives in the next generation.	>1		5
Evolutionary Algorithms	`parents_structure_fraction`	numb	The fraction of individuals in the current population that is used to generate structures in the next cycle.	0-1		0.75
Evolutionary Algorithms	`mutation_fraction`	numb	The fraction of individuals in the population that will undergo mutation in each generation.	0-1		0.2
Evolutionary Algorithms	`heredity_fraction`	numb	The fraction of individuals in the population that will be generated through heredity (crossover/recombination) operations between two or more parents.	0-1		0.6
Evolutionary Algorithms	`permutation_fraction`	numb	The fraction of individuals in the population that will undergo a permutation operation (e.g.,	0-1		0.1

Category	Data Field	Type	Definition	Constraints	Units	Example
			swapping atom positions within a structure) in each generation.			

2.3 Particle Swarm Optimisation Algorithms

Category `_csp.particle_swarm_optimisation_[]` : Subgroup for CSP Structure Generation methods that use Particle Swarm Optimisation. For these fields to be used, the `_csp.structure_generation.method` must include "Particle Swarm Optimisation".

Category	Data Field	Type	Definition	Constraints	Units	Example
Particle Swarm Optimisation	`population_size`	numb	The number of candidate crystal structures (particles) in the swarm.	>0		50
Particle Swarm Optimisation	`number_of_generations`	numb	The maximum number of optimization cycles (generations or iterations) the PSO algorithm will run.	>0		100
Particle Swarm Optimisation	`inertia_weight`	numb	A parameter controlling the contribution of the previous velocity of the particle to its current velocity.	0-1		0.7
Particle Swarm Optimisation	`max_inertia_weight`	numb	If the inertia weight changes with each iteration, this parameter specify the maximum value it can have.	0-1		0.9
Particle Swarm Optimisation	`min_inertia_weight`	numb	If the inertia weight changes with each iteration, this parameter specify the minimum value it can have.	0-1		0.4
Particle Swarm Optimisation	`cognitive_coefficient`	numb	A parameter (also called self-confidence factor) controlling the influence of the particle's own best position found so far on its movement.	>=0		2
Particle Swarm Optimisation	`social_coefficient`	numb	A parameter (also called swarm confidence factor) controlling the influence of the swarm's best position found so far on the particle's movement.	>=0		2
Particle Swarm Optimisation	`velocity_clamp_max`	numb	The maximum allowed velocity for each dimension if velocity clamping is enabled.	>0		0.2

2.4 Simulated Annealing

Category `_csp.simulated_annealing_[]` : Subgroup for CSP Structure Generation methods that use Simulated Annealing. For these fields to be used, the `_csp.structure_generation.method` must include "Simulated Annealing".

Category	Data Field	Type	Definition	Constraints	Units	Example
Simulated Annealing	`initial_temperature`	numb	The starting temperature of the simulated annealing process.	>0	K	500
Simulated Annealing	`cooling_rate`	numb	The parameter that determine how the temperature is decreased over the course of the simulation.	0-1		0.95
Simulated Annealing	`number_of_steps`	numb	The number of attempted structure generation and acceptance steps performed at each temperature.	>0		10

2.5 Monte Carlo Parallel Tempering

Category `_csp.monte_carlo_parallel_tempering_[]` : Subgroup for CSP Structure Generation methods that use Monte Carlo Parallel tempering. For these fields to be used, the `_csp.structure_generation.method` must be set to "Monte Carlo Parallel Tempering".

Category	Data Field	Type	Definition	Constraints	Units	Example
Monte Carlo Parallel Tempering	`number_of_replicas`	numb	The number of independent Monte Carlo simulations (replicas) running in parallel at different temperatures.	1		3
Monte Carlo Parallel Tempering	`temperatures_list`	list	The list of temperatures at which the replicas are run.	`[T >= 0]`	K	`[0, 300, 600]`

Category	Data Field	Type	Definition	Constraints	Units	Example
Monte Carlo Parallel Tempering	`number_of_steps`	numb	The number of Monte Carlo steps performed by each replica at its assigned temperature in each parallel tempering cycle.	>0		100

2.6 Random Search

Category `_csp.random_[]` : Subgroup for CSP Structure Generation methods that use Random, Quasi-random algorithms. For these fields to be used, the `_csp.structure_generation.method` should be set to "Random Sampling".

Category	Data Field	Type	Definition	Constraints	Units	Example
Random Search	`random_numbers_algorithm`	char	Specifies the type of random algorithm used.	"Pseudorandom", "Quasirandom", "Other"		"Pseudorandom"
Random Search	`number_of_samples`	numb	The total number of unique crystal structures to be generated and evaluated during the random search.	>0		5000

Examples

Search in all space groups after 100 structures are generated with an evolutionary algorithm:

```
_csp.structure_generation.space_group_number_list "all"
_csp.structure_generation.method "Evolutionary Algorithm"
_csp.structure_generation.density_lower_limit 750
_csp.structure_generation.density_upper_limit 1500
_csp.structure_generation.stopping_criteria "Max Structures"
_csp.structure_generation.stopping_criteria_max_structures_evaluated 100
```

Combination of different structure generation methods and on most popular space groups for organic crystals:

```
_csp.structure_generation.space_group_number_list [14 2 15 61 19 4 33 29 5 1]
_csp.structure_generation.density_lower_limit 750
_csp.structure_generation.density_upper_limit 1500
_csp.structure_generation.method ["Random Sampling" "Simulated Annealing"]

# Random Search
_csp.random.random_numbers_algorithm "Quasi-random"
_csp.random.number_of_samples 50

# Simulated Annealing
_csp.simulated_annealing.initial_temperature 400
_csp.simulated_annealing.cooling_rate 0.95
_csp.simulated_annealing.number_of_steps 100
```

3. Structure Ranking Methods (High-level)

Within this section, you can define the workflow used to rank the different crystals and give high-level details of the methods used.

3.1 General Fields

Category `_csp.structure_ranking_[]` : Category for structure ranking and optimisation methods.

Category	Data Field	Type	Definition	Constraints	Units	Example
Structure Ranking	`method`	char	The energy or scoring model used to rank structures.	- pDFT (Sec. 3.2) - Forcefield (Sec. 3.3) - Semi-Empirical (Sec. 3.4) - Wavefunction (Sec. 3.5) - ML Potentials (Sec. 3.6) - Other		Forcefield
Structure Ranking	`calculation_type`	char	Indicates how atomic positions are changed.	- Optimisation - Ensemble Average - Single point		"Single point"

Category	Data Field	Type	Definition	Constraints	Units	Example
Structure Ranking	`software_citation`	char	Details of the software used for structure generation.	Free Text		
Structure Ranking	`software_version`	char	Version of software used for structure generation.	Free Text		
Structure Ranking	`stage`	numb	In case of multi-step approaches, the stage of the ranking method.	>=0		0
Structure Ranking	`stage_id`	char	In case of multi-step approaches, the stage identifier of the ranking method.	Free Text		FF, PBE, PBE0
Geometry Optimisation	`algorithm`	char	Geometry optimisation algorithm	- BFGS - L-BFGS - Quasi-Newton - FIRE - Steepest Descent - Conjugate Gradient - Other		
Geometry Optimisation	`cell`	char	It can be "fixed" for no cell optimisation, "isotropic" or "anisotropic" for cell relaxation calculations.	- fixed - isotropic - anisotropic		
Geometry Optimisation	`atoms`	char	It can be "fixed" for no atoms' position optimisation, "all" for all-atoms geometry optimization, "hydrogens" for optimisation of only H atoms, "non-hydrogens" for non-H atoms or a list of atoms for custom relaxation.	- fixed - all - hydrogens - non-hydrogens - `[List of _chemical.conn_atom.number]`		`[1 2 3 4]`
Geometry Optimisation	`relax_force_convergence`	numb	Convergence criteria for stopping the geometry optimisation. Present in TCOD as `_dft_atom_relax_force_conv`.	>0	kJ mol^{-1} nm^{-1}	0.1

Examples

3.2 Periodic Density Functional Theory

Category `_dft_[]` : Subgroup for CSP Structure Ranking methods that use pDFT methods (the *p* of *pDFT* is removed in `_dft` for consistency with the TCOD Dictionary). For these fields to be used, the `_csp.ranking_method` should be set to "pDFT".

Category	Data Field	Type	Definition	Constraints	Units	Example
pDFT	`exchange_correlation_functional_type`	char	Specifies the type of exchange-correlation functional used.	- LDA - GGA - meta-GGA - Hybrid - Other		GGA
pDFT	`exchange_correlation_functional_name`	char	Specifies the name of exchange-correlation functional used.	- PBE - PBE0 - SCAN - ...		PBE
pDFT	`pseudopotential_type`	char	Defines the type of pseudopotentials used.	- Plane-waves - PAW - Norm-conserving - Ultrasoft		PAW
pDFT	`dispersion_correction`	char	The Van der Waals correction used.	- Grimme-D2 - Grimme-D3 - Tkatchenko-Scheffler		XDM

Category	Data Field	Type	Definition	Constraints	Units	Example
				- Many-body dipersion - XDM - Other		

3.3 Forcefields

Category `_forcefield_[]` : Subgroup for CSP Structure Ranking methods that use forcefield or mixed inter/intra molecular methods. For these fields to be used, the `_csp.ranking_method` should be set to "Forcefield".

Category	Data Field	Type	Definition	Constraints	Units	Example
Forcefield	`name`	char	Name of the force field.	Free Text		
Forcefield	`intramolecular_term`	char	The energy evaluation method for intramolecular interactions.	- "Bonded Parameters" - "Isolated Molecule Energy" - None - Other		
Forcefield	`electrostatic_term`	char	Functional form of electrostatic interactions	- "Point-Charges" - Multipoles - Other		
Forcefield	`vdw_term`	char	Functional form of van der Waals interactions	- LJ(C6 -C12) - LJ(epsilon -sigma) - Buckingham - ReaxFF Morse-Potential - 14-7 function - Other		
Forcefield	`parameterization_method`	char	Briefly describes the primary method used to derive the force field parameters.	Free Text		"Fitting to gas-phase QM data", "Transferable parameters based on atom types"
Forcefield	`qm_parameterization_functional`	char	The exchange-correlation functional used in the gas-phase quantum mechanical calculations when fitting force field parameters.	- "MP2" - "CCSD(T)" - "B3LYP" - ...		MP2
Forcefield	`qm_parameterization_basis_set`	char	The basis set used in the gas-phase quantum mechanical calculations when fitting force field parameters.	- "aug-cc-pVTZ" - "6-31G(d,p)" - ...		6-31G

3.4 Semi-Empirical

Category `_semiempirical_[]` : Subgroup for CSP Structure Ranking methods that use Semi-Empirical methods. For these fields to be used, the `_csp.ranking_method` should be set to "Semi-Empirical".

Category	Data Field	Type	Definition	Constraints	Units	Example
Semi-Empirical	`method`	char	Specifies the name of the Semi-Empirical method used.	- AM1 - PM3 - PM6 - xTB - ...		PM6

Category	Data Field	Type	Definition	Constraints	Units	Example
Semi-Empirical	`electronic_parameters`	char	The Slater-Koster tables or equivalent defining the atomic orbitals and pairwise element-element interactions	- mio - 3ob - ...		mio
Semi-Empirical	`repulsive_potential`	char	When not included in the SK files, the repulsive potential term used.			
Semi-Empirical	`hydrogen_bond_correction`	char	H-bond corrections for semi-empirical methods.	- H+ - H4 - ...		H+
Semi-Empirical	`dispersion_correction`	char	Dispersion corrections for semi-empirical methods	- D3 - TS - MBD - ...		D3

3.5 Wavefunction

Category `_wavefunction_[]` : Subgroup for CSP Structure Ranking methods that use wavefunction methods. For these fields to be used, the `_csp.ranking_method` should be set to "Wavefunction".

Category	Data Field	Type	Definition	Constraints	Units	Example
Wavefunction	`exchange_correlation_functional`	char	Specifies the name of functional used.	- HF - MP2 - CC - ...		MP2
Wavefunction	`basis_set_type`	char	Defines the type of basis used.	- GTH - NAO - ...		NAO

3.6 ML Potentials

Category `_ml_potential_[]` : Subgroup for CSP Structure Ranking methods that use machine learning potentials methods. For these fields to be used, the `_csp.ranking_method` should be set to "ML Potentials".

Category	Data Field	Type	Definition	Constraints	Units	Example
ML Potential	`method`	char	Specifies the name of the ML Potential used. In case of ML parameterisation of classical forcefields, refer to the Forcefields dictionaries	- ANI - Mace - GAP/SOAP - ...		

3.7 Free Energy

Category `_free_energy_[]` : Subgroup for CSP Structure Ranking methods that use free energy methods.

Category	Data Field	Type	Definition	Constraints	Units	Example
Free Energy	`method`	char	Specifies the name of the approach used to calculate free energies.	- HA - QHA - PSCP - EC - ...		QHA
Free Energy	`reference_temperature`	numb/list	The temperature or list of temperatures at which free energies are calculated	>0 or `List[PositiveFloat]`	K	`[ 100.0, 200.0, 300.0 ]`
Free Energy	`reference_pressure`	numb/list	The pressure or list of pressures at which free energies are calculated	>0 or `List[PositiveFloat]`	Pa	`[ 100000.0, 200000.0 ]`

Examples

pDFT with hybrid XC functional and additional datafields taken from the TCOD DFT dictionary:

```
_csp.structure_ranking.method DFT
_csp.structure_ranking.stage_id final

_dft.exchange_correlation_functional_type GGA
_dft.exchange_correlation_functional_name PBE
_dft.pseudopotential_type PAW
_dft.dispersion_correction XDM

_dft.kinetic_energy_cutoff_wavefunctions 600
_dft.atom_relax_force_conv 0.002
_dft.BZ_integration.method "Monkhorst-Pack"
_dft.BZ_integration.grid_dens_X 0.5
_dft.BZ_integration.grid_dens_Y 0.5
_dft.BZ_integration.grid_dens_Z 0.5
```

Multiple energy evaluation steps:

```
data_method
_chemical.name urea
_csp.structure_generation.space_group_number_list all
_csp.structure_generation.method "Particle Swarm Optimisation"

loop_
    _csp.structure_ranking.stage
    _csp.structure_ranking.stage_id
    _csp.structure_ranking.calculation_type
    _geometry_optimisation.atoms
    _geometry_optimisation.cell
    _csp.structure_ranking.method
    0 "gaff"    "Optimisation" "all" "anisotropic" "Forcefield"
    1 "psi_mol" "Optimisation" "all" "anisotropic" "Forcefield"
    2 "pbe"     "Optimisation" "all" "anisotropic" "DFT"
    3 "pbe0"    "Single-Point" .     .             "DFT"

# END

# Additional Parameters for each method
data_gaff
_ff.name "GAFF"
_ff.intramolecular_term "Bonded-Parameters"
_ff.electrostatic_term "Point-Charges"
_ff.vdw_term "LJ(epsilon,sigma)"
_ff.parameterization_method "BCC"
_ff.qm_parameterization_functional "AM1"
# END

data_psi_mol
_ff.name "Psi_mol"
_ff.intramolecular_term "Isolated Molecule Energy"
_ff.electrostatic_term "Multipoles"
_ff.vdw_term "Buckingham"
_ff.parameterization_method "GDMA"
_ff.qm_parameterization_functional "PBE0"
_ff.qm_parameterization_basis_set "6-31G(d,p)"
# END

data_pbe
_dft.exchange_correlation_functional_type "GGA"
_dft.exchange_correlation_functional_name "PBE"
# END

data_pbe0
_dft.exchange_correlation_functional_type "Hybrid"
_dft.exchange_correlation_functional_name "PBE0"
_dft.dispersion_correction "TS"
# END
```

4. Predicted Crystal Structure

Describes the structure-specific outputs of CSP methods. Category `_predicted_structure_[]`

Category	Data Field	Type	Definition	Constraints	Units	Example
Predicted Structure	`temperature`	numb	The temperature at which the energy and other properties of the theoretical structure were calculated.	>=0	K	298.15

Category	Data Field	Type	Definition	Constraints	Units	Example
Predicted Structure	`pressure`	numb	The pressure at which the energy and other properties of the theoretical structure were calculated.	:	Pa	101325.0
Predicted Structure	`calculated_density`	numb	The calculated density of the crystal	>=0	kg m^{-1}	1420.0
Predicted Structure	`total_energy`	numb	The total energy of the theoretical structure, i.e. energy relative to all of the nuclei and electrons seperated to an infinite distance.	:	kJ mol^{-1}	-1500.5
Predicted Structure	`absolute_lattice_energy`	numb	The absolute lattice energy of the crystal, i.e. energy relative to all the molecules seperated to an infinite distance.	:	kJ mol^{-1}	-1600.8
Predicted Structure	`absolute_free_energy`	numb	The absolute free energy of the crystal.	:	kJ mol^{-1}	-1450.2
Predicted Structure	`free_energy_correction`	numb	The correction applied to the lattice energy to obtain the free energy, accounting for vibrational and other thermal effects.	:	kJ mol^{-1}	50.6
Predicted Structure	`relative_lattice_energy`	numb	The lattice energy of the theoretical structure relative to the lowest energy structure found in the CSP.	>=0	kJ mol^{-1}	0.0, 5.2
Predicted Structure	`energy_uncertainty`	numb	An estimate of the uncertainty associated with the calculated energy of the theoretical structure.	>=0	kJ mol^{-1}	0.1
Predicted Structure	`score`	numb	To allow for methods that may rank by criteria other than energies (e.g., based on stability or other desired properties).	:		1, 0.3333, 0.01
Predicted Structure	`rank`	numb	The rank of the structure when ordered by chosen criteria where 1 is considered to be the most favorable or likely structure.	>=1		2, 7, 12

Details on composition, unit cell, symmetry, and atomic coordinates can be specified through the CIF Core dictionary.

Examples

```
data_A_1
# Structure is thoeretically generated
_exptl.method 'theoretical model'

# Ranking Step
_csp.structure_ranking.stage 1
_csp.structure_ranking.stage_id dftb

# Properties
_predicted_structure.temperature 0
_predicted_structure.relative_lattice_energy 1.5
_predicted_structure.rank 5

# Crystal
_symmetry.cell_setting           monoclinic
_symmetry.space_group_name_H-M   'P 21/c'
_symmetry.Int_Tables_number      14
_space_group_name_Hall           '-P 2ybc'
loop_
_symmetry.equiv_pos_site_id
_symmetry.equiv_pos_as_xyz
1 x,y,z
2 -x,1/2+y,1/2-z
3 -x,-y,-z
4 x,1/2-y,1/2+z
_cell.length_a                   16.8168
_cell.length_b                   7.0178
_cell.length_c                   14.5475
_cell.angle_alpha                90
_cell.angle_beta                 115.28
_cell.angle_gamma                90
_cell.volume                     1552.44
loop_
_atom_site.label
_atom_site.type_symbol
```

```
_atom_site.fract_x
_atom_site.fract_y
_atom_site.fract_z
C_00 C 0.402802 0.483043 0.842739
C_01 C 0.387973 0.677363 0.792505
H_02 H 0.372266 0.3709 0.785899
C_03 C 0.290934 0.723546 0.722019
C_04 C 0.299744 0.784092 0.627602
O_05 O 0.430454 0.669327 0.717154
N_06 N 0.375029 0.749323 0.62653
C_07 C 0.433253 0.840791 0.865082
H_08 H 0.373059 0.480692 0.896797
H_09 H 0.473281 0.453133 0.883342
H_0a H 0.406023 0.852375 0.921481
H_0b H 0.422667 0.976203 0.823613
H_0c H 0.504341 0.815441 0.904886
H_0d H 0.247676 0.59825 0.706716
H_0e H 0.264985 0.835703 0.754024
S_0f S 0.217342 0.894753 0.52042
O_0g O 0.132826 0.836886 0.515882
O_0h O 0.235781 0.880775 0.432108
C_0i C 0.22914 1.15888 0.556705
F_0j F 0.218005 1.16897 0.646067
F_0k F 0.314654 1.21064 0.581184
C_0l C 0.161031 1.27054 0.472953
C_0m C 0.174831 1.34683 0.391627
C_0n C 0.0755211 1.2804 0.468012
C_0o C 0.00748411 1.36477 0.384876
F_0p F 0.256854 1.33974 0.393375
C_0q C 0.107309 1.42929 0.307776
C_0r C 0.0231417 1.4367 0.304553
H_0s H -0.0306764 1.49657 0.238225
H_0t H -0.0580026 1.37054 0.382481
H_0u H 0.062589 1.21934 0.529056
H_0v H 0.120713 1.48531 0.246093
```

5. Conventions

A few guidelines are adopted in the description of specific data fields as highlighted in the table below. Except for *pDFT*, full names are preferred.

Category	Data Field	Suggested Input Item	Alternatives to avoid
Structure Generation	`method`	Random Sampling	Quasirandom, Pseudorandom (specified in separate datafield)
Structure Generation	`method`	Evolutionary Algorithm	Genetic Algorithm, EA, GA
Structure Ranking	`method`	Forcefield	Force Field, Force-Field, FF
Structure Ranking	`method`	pDFT	DFT, Density Functional Theory, periodic-DFT
Structure Ranking	`method`	Semi-Empirical	Semi Empirical

In addition, the Structure Ranking `method` "ML Potential" refers to methods using *ad hoc* descriptors for neural network training to directly compute energy and forces. On the other hand, ML models used to parameterise models constants should be classified in the related method. For example, forcefield constants parametrised with a ML network should be classified as "Forcefield".

6. Future Developments

A few areas relevant to CSP have not been explored yet and might be included in later updates of the dictionary. In general, new or specific methods can use the "Other" option and specify possible publications describing the workflow.

A list of missing sections is shown below:

- Initial molecule (or list of molecules and conformers) coordinates and properties.
- ML-based Structure Generation methods.
- Clustering algorithms used to remove duplicates.
- While the TCOD dictionary is available for DFT methods and a draft dictionary for forcefield methods is being developed, datafields of other energy evaluation methods are limited to basic identification labels. This includes:
  - Semi-Empirical methods
  - ML Potentials
  - Free energy correction methods
- Output structure properties are limited to the energy or score of the crystal. Other measurable properties (the band gap for example) are not currently included.