

Architector for high-throughput cross-periodic table 3D complex building

Received: 7 July 2022

Accepted: 18 April 2023

Published online: 15 May 2023

Michael G. Taylor^{1,2}, Daniel J. Burrill^{1,2}, Jan Janssen¹, Enrique R. Batista¹, Danny Perez¹ & Ping Yang¹

Rare-earth and actinide complexes are critical for a wealth of clean-energy applications. Three-dimensional (3D) structural generation and prediction for these organometallic systems remains a challenge, limiting opportunities for computational chemical discovery. Here, we introduce *Architector*, a high-throughput in-silico synthesis code for s-, p-, d-, and f-block mononuclear organometallic complexes capable of capturing nearly the full diversity of the known experimental chemical space. Beyond known chemical space, *Architector* performs *in-silico* design of new complexes including any chemically accessible metal-ligand combinations. *Architector* leverages metal-center symmetry, interatomic force fields, and tight binding methods to build many possible 3D conformers from minimal 2D inputs including metal oxidation and spin state. Over a set of more than 6,000 x-ray diffraction (XRD)-determined complexes spanning the periodic table, we demonstrate quantitative agreement between *Architector*-predicted and experimentally observed structures. Further, we demonstrate out-of-the box conformer generation and energetic rankings of non-minimum energy conformers produced from *Architector*, which are critical for exploring potential energy surfaces and training force fields. Overall, *Architector* represents a transformative step towards cross-periodic table computational design of metal complex chemistry.

Data-driven methods for materials¹⁻³ and chemical^{4,5} discovery have enjoyed considerable success, fueled by the availability of accurate high-throughput structure-based simulation approaches⁶⁻⁹. For example, the dramatic increase in availability of computational data for d-block organometallics was in part enabled by the creation of reliable three-dimensional (3D) structure generation tools; notable examples being the *molSimplify*⁶, *molAssembler*⁷, and *DENOPTIM*⁸ codes. These tools allow for the generation of 3D configurations of organometallic complexes, whose properties can subsequently be computed using high-level electronic structure calculations. *molSimplify* structure generation operates using force-field pre-optimized ligands and assembly of those ligands through alignment to a specified (typically octahedral) metal-center geometry. *molAssembler* operates by leveraging graph enumeration⁹, stereopermuters, and the distance

geometry¹⁰ algorithm informed by tabulated bond lengths. Another approach for organometallic structure generation is taken by the *DENOPTIM*⁸ program, which leverages fragment building in addition to genetic algorithms to assemble hypothetical complexes that optimize a given fitness function. The generation of large numbers of potential complexes, via tools such as the ones listed above, can fuel the computational creation of massive datasets of chemical properties. This, in turn, can aid the development of powerful data-driven machine-learning approaches for the efficient exploration of chemical space.

A complex building algorithm to set up initial structures of lanthanoid complexes has been published by Munguba and co-workers¹¹. It considers stereo control, including stereoisomer identification and coordination chirality recognition, and was demonstrated for monodentate and bidentate ligands. f-block organometallic chemistry is vital

¹Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ²These authors contributed equally: Michael G. Taylor, Daniel J. Burrill.

✉ e-mail: erb@lanl.gov; danny_perez@lanl.gov; pyang@lanl.gov

for clean energy and the nuclear energy cycle, including chemical separations for waste treatment^{12,13}, as well as the extraction of rare earth elements¹⁴ needed for modern technologies ranging from cell phones to turbines¹⁵. The relative rarity of data on f-block systems can be explained by a range of fundamental and practical factors, including difficulties inherent to working with radioactive elements, elemental scarcity^{15,16}, complexity in terms of characterization^{17–19}, high coordination environment, and their complex electronic structure^{20,21}. Nonetheless, there are noteworthy cases where computational design of organometallics has been successfully performed^{22,23}, albeit with structures intuitively-generated by hand and with ligand sets on the order of tens of ligands rather than hundreds or thousands. With these methods, however, large-scale chemistry and structure generation including conformers for potential f-block organometallic complexes remains prohibitively time-consuming, limiting the potential application of large-scale data-driven methods for f-block organometallic chemistries. Because of these challenges, the amount of data on the chemical structure of f-block organometallics is dwarfed by data pertaining to related d-block organometallics^{24–26}.

Older 3D-generation packages were benchmarked with root-mean-square displacement (RMSD) comparisons to experimental structures²⁷ while recent 3D structure generation routines have not compared generated structures to experimental structures either due to interests more targeted towards specific application spaces^{6,8}, or due to concerns surrounding metrics for comparison⁷. Notably, gas phase or liquid phase conformers can dramatically differ from crystalized structures²⁸, but different conformers can be vital for accurately capturing spectroscopic²⁹ and even catalytic³⁰ properties of the underlying complex. Key advances for conformer sampling and the generation and prediction of thousands of structures are the broadly applicable semi-empirical extended tight binding (*xTB*) *xTB*/GFN2-*xTB*³¹ and related geometry relaxation routines³². Both were leveraged to produce the largest available electronic structure-based d-block dataset derived directly from experimental structures²⁵. Further methods for enhanced conformer sampling beginning from existing 3D structures show promise for d-block applications³³. Studies of d-block conformers, generated from a limited set of CSD geometries (~40 structures), revealed lower-energy structures for 68% of CSD structures without any shifts in relative ligand positioning around the metal center³⁴. However, with multiple ligands and different symmetries, additional lower-energy solution-phase conformers for even these simpler d-block systems exist, much less with higher-coordinated f-block systems³⁵. The availability of large databases of electronic and chemical structures of different conformers and off-equilibrium structures of f-block complexes could drive a more systematic exploration of this vast chemical space. Such a database would also enable the development of powerful accelerated computational methods such as tight binding^{36,37} and machine-learning (ML) potentials^{38,39} capable of capturing f-block long timescale physicochemical properties such as diffusion coefficients and solution-phase reaction rates in f-block extraction processes.

To address the challenges inherent in f-block organometallic structure generation and their relative data paucity, we here introduce *Architector*, a python package for mononuclear organometallic 3D conformer assembly from 2D-inputs based on metal-center symmetry analysis, distance geometry, fragment assembly, and *xTB*-evaluation. To sensibly measure deviations between generated structures where metal-distal configurations overaccentuate differences, we tailor a simple RMSD approach limited to metal-center proximal alignment. Leveraging this RMSD method and energetic comparisons, we show *Architector* can reproduce X-ray diffraction (XRD)-determined 3D chemical structures from the Cambridge Structural Database²⁶ (CSD) for f-block in addition to s-, d-, and p- block chemistries. We further show how leveraging tight binding methods inside the *Architector* generation workflow allows for near-density functional theory (DFT) levels of accuracy in ordering generated conformers out of the box.

Finally, we highlight its high-throughput capabilities, producing up to 20 conformers evaluated with *xTB*/GFN2-*xTB* for each of over 6,000 CSD structures within 12 h on ~500 cores.

Results

Architector workflow and design

We begin with an overview of *Architector* structure generation along with example chemical illustrations of each step of the process (Fig. 1). We showcase a Ce complex for example inputs (Fig. 1a) and outputs (Fig. 1e) due to its variety of multidentate ligands (both bidentate cis and tridentate fac). The example *Architector* inputs are derived from an existing molecule by identifying the metal, coordination number (CN), ligand simplified molecular-input line-entry system (SMILES) strings⁴⁰, and which index/indices from the SMILES are coordinating atoms (CA) to the metal center. *Architector* includes a utility to aid users in identifying CAs for arbitrary ligand SMILES (Supplementary Fig. 1). In total, *Architector* inputs amount to a full 2D molecular graph specification for the 3D molecule to be constructed.

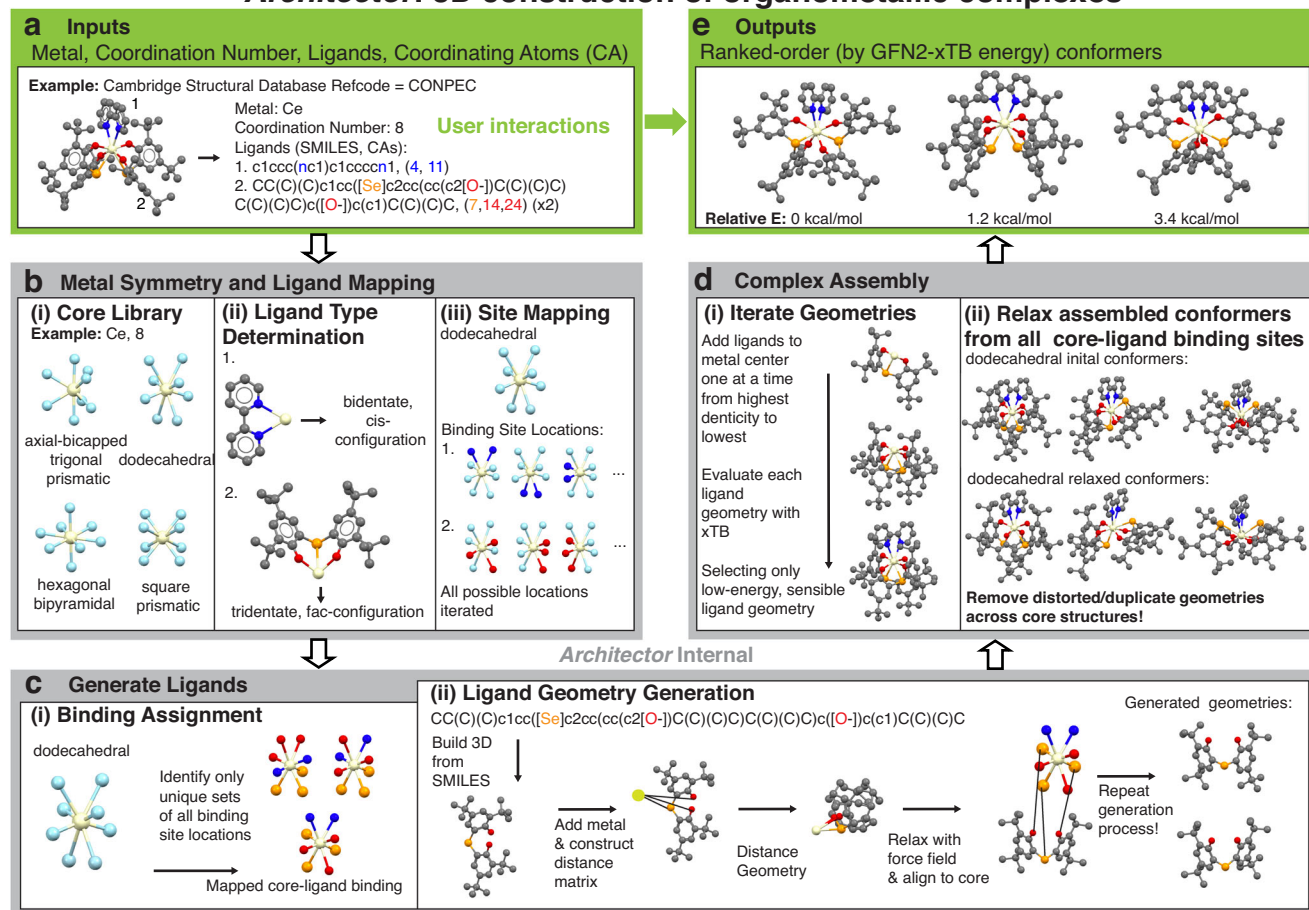
2D molecular graphs can be defined for any synthesized complex and any not-yet-synthesized complex from across the periodic table for *Architector* structure generation. For user ease in generating 2D molecular graphs and assigning electronic states to complexes, *Architector* also contains default information for each metal including oxidation states, spin, and coordination numbers in addition to a substantial and increasing number (~100) of named ligands that can be used in construction (Supplementary Table 5). After the *Architector* 3D construction, the user is returned a list of generated conformers (Fig. 1e) for the given input, which is ranked according, by default, to their GFN2-*xTB*³¹ energy.

As deciding if a 2D molecular graph can be embedded into 3D is NP-hard⁴¹, we opt for heuristic approaches. These approaches attempt to identify close-to-minima energy structures as well as a range of reasonable higher-energy structural isomers that can be useful for understanding high-temperature solution chemistries and for training sets for reduced-cost semi-empirical approaches³⁷. For the given metal type and CN, all pre-defined core geometries (Supplementary Table 1) are referenced and tested (Fig. 1b,i). As with the default core library, we identified default ligand types and corresponding CA-M-CA angles (Fig. 1b,ii) from ligands sampled across the CSD (Supplementary Note 1, Supplementary Fig. 2, and Supplementary Table 2). Finally, with the given ligand and coordinating atoms, ligand types can be assigned using an included aid (Supplementary Fig. 3) or from a built-in brute-force search method. Thus, both ligand and core geometries are assigned or determined by *Architector*.

Given the ligand and core geometries, *Architector* constructs all possible mappings between the two. *Architector* determines all ligand binding sites by taking all combinations of core CA positions and determining if the angles between all pairs of CAs align with the assigned ligand geometry (Fig. 1b,iii and Supplementary Fig. 4). Then, *Architector* assigns valid sets of binding sites from all possible binding sites for all ligands by reducing to only sets of possible binding sites with no shared core CAs (Fig. 1c,i).

Even for simple complexes, the number of possible core-ligand binding site mappings can be enormously large, needing that one only tests a reduced number of mappings. For example, a CN = 8 complex bound by 8 different monodentate ligands has $8! = 40,320$ possible core-ligand binding site maps. However, if some of the ligands are identical (e.g., 4 water and 4 ethanol ligands), the number of unique mappings usually decreases by 2–3 orders of magnitude (8 choose 4 = 70). Thus, for cases with fewer distinct ligands, mappings can automatically be greatly reduced.

Mappings can be further reduced by not considering structures with identical metal-center symmetries. To evaluate the uniqueness of valid sets of binding sites and quickly rank the likelihood of each mapped core-ligand binding, we developed a heuristic function corresponding to a pseudo-energy of binding sites (Supplementary Note 2).

Architector: 3D construction of organometallic complexes**Fig. 1 | Flow chart of Architector: 3D construction of organometallic complexes.**

a Example inputs for an Architector generation derived from a Cambridge Structural Database (CSD) molecule **b** Metal-center symmetry and ligand mapping procedures done in order (i)-(iii), steps (ii) and (iii) are repeated for each core geometry. **c** Ligand binding site assignment mapping (repeated for each core geometry) followed by ligand geometry generation for each binding site mapping. **d** Full complex conformer assembly and evaluation performed for each core

geometry and unique binding site mapping. **e** Outputs from the Architector generation including multiple conformers sorted by GFN2-xTB energy. Note that users interact directly with only **a** and **e** while **b**, **c**, **d**, and **e** are handled internally by Architector. Inset images are molecules colored by atom type with N:blue, Ce:yellow, C:gray, O:red, Se:orange, and core geometry binding site:light blue, hydrogens have been removed for clarity.

We note that the pseudo-energy evaluated with $b=1$ heuristic charge produced -94% of GFN2-xTB lowest-energy conformers considering only the top 10 lowest-pseudo-energy binding configurations (Supplementary Fig. 5), while using up to the top 19 lowest-pseudo-energy configurations produced 99.5% of lowest-energy conformers. Therefore, pre-calculating pseudo-energy greatly reduces the number of sets of binding sites that must be considered if only minimum-energy structures are desired.

Architector 3D generation of structures then starts from ligands and assigned ligand CA binding sites (Fig. 1c,ii). We developed the ligand 3D-generation routines extending and synergizing approaches from both *molSimplify*⁶ and *molAssembler*⁷. First, we initialize chemically-meaningful ligand geometries from the SMILES strings using the *Openbabel*⁴² package, which are then relaxed with either MMFF94⁴³ or UFF⁴⁴. Next, we prepare ligands for distance geometry⁴⁵ conformer generation⁴⁶. Briefly, we form both lower-bound and upper bound distances based on the initial ligand geometries in addition to single bond covalent radii (r_{cov})⁴⁷, and van der waals radii⁴⁸ (vdw) found in the literature. Additional details surrounding distance matrix construction and distance geometry generation can be found in Supplementary Table 3 and Supplementary Note 3. To correct the distortions resulting from distance geometry, we turn again to FFs with a multistep geometry cleaning process

(Supplementary Note 4). Following FF cleaning, the Kabsch algorithm⁴⁹ is used to rotate the ligand around the metal to minimize CA location deviation from the assigned core geometry locations, returning the rotational root-mean-squared deviation (rotRMSD) between the sets of locations (Supplementary Note 5). After force-field relaxation and alignment, the generated ligand conformers are ready for metal-ligand complex assembly.

With different ligand conformations at multiple sites at the metal surface, rapid methods for ranking conformations and reducing numbers of structures breaking chemical sanity are needed for complex assembly. By default we use GFN2-xTB³¹ to select FF-relaxed ligand geometries by placing each ligand conformer on the complex and evaluating the total energy to select low-total-energy geometries for each ligand in order from highest to lowest denticity (Fig. 1d,i). To bias the conformers to their assigned binding sites, we select the conformers with the lowest total GFN2-xTB energy multiplied by $1/\text{rotRMSD}$ to minimize potential ligand overlaps. As this bias does not remove all potential ligand overlaps, we introduce three distinct interatomic distance cutoff checks before GFN2-xTB evaluation to minimize tests on unreasonable geometries (Supplementary Note 6, Supplementary Fig. 6, and Supplementary Table 4). If different methods for ordering of ligands during construction are desired, users can optionally request all the assembled complexes and specify any other

desired ordering (Supplementary Fig. 7). Assembled complexes are then ready for relaxation and electronic structure analysis.

Final relaxation on assembled complexes can be performed with any electronic structure or force-field method, though by default *Architector* uses GFN2-xTB. If unspecified by the user, *Architector* assigns default oxidation state, spin state, and core CNs (Supplementary Table 5) to the complex based on the metal identity and total electron counts (Supplementary Note 7). For the actinides, we swap the metal center and electron counts for the equivalent-group lanthanide, since atomic numbers greater than 88 are not supported by GFN2-xTB³¹. However, this swap is unnecessary for models parametrized for actinides³⁷. With charge and spin assigned, we relax all assembled complexes within the Atomic Simulation Environment (ASE)⁵⁰. Once relaxed, we perform an additional set of tighter interatomic distance chemical sanity checks ensuring generated structures match the input 2D molecular graph (Supplementary Table 4). For all relaxed complexes found to satisfy the input parameters, *Architector* by default applies Kuhn–Munkres ordered RMSD (kmRMSD)⁵¹ to eliminate duplicate structures before returning all generated complexes ranked by GFN2-xTB energy (Fig. 1e). Additional comments surrounding assumptions made during complex construction can be found in Supplementary Note 8. From input to output structures, *Architector* is thus uniquely tailored to high-throughput and accurate organometallic complex 3D structure generation.

Structure reference for targeted generation and CSD-replication set

To create a targeted set of mononuclear complexes for 3D structure generation, we turned to the CSD²⁶. We mined the CSD version 5.43 (Nov 2021) for mononuclear structures (Fig. 2a and Supplementary Note 8) using previously described methods⁵², resulting in a total of 312,527 mononuclear complexes. The full CSD distributions of mononuclear metal centers are 83% d-block with 80% CN less than 8 and primarily coordinated by C, N, or O atoms (Supplementary Fig. 8). From here, the *CSD python API* was used to remove structures incompatible with *Architector*, broadly due to mismatches in stoichiometry between the 2D representation and 3D structures, yet-unsupported ligands, unreasonable charges, or structures with distorted geometries (Supplementary Table 6). We highlight example *Architector*-compatible structures from CSD (Fig. 2b) that are representative of different blocks, visually noting that f-block complexes tend to be higher-coordinated. To focus on a more evenly distributed set of complexes across metal (some metals have fewer than ten structures), we randomly sampled up to 100 different complexes for

each metal resulting in a subset of 6,154 structures for generation (Fig. 2c and Supplementary Table 7). The atom distributions and descriptive statistics for this subsampled set are still largely representative of the full CSD (Supplementary Fig. 8 and Supplementary Fig. 9). Thus, we created a cross-periodic table set of organometallic complexes for replication by *Architector*.

Structure and energy statistics for lowest-energy *Architector* conformers vs. CSD

Across the full set of 6,154 structures, we performed *Architector* generation using CSD-derived oxidation state assignments (Supplementary Note 7) and corresponding spin states suggested by the *mendeleev*⁵³ package and relaxed the CSD-derived structures with GFN2-xTB. Overall, this resulted in 5,956 successful generation of sane relaxed complexes with comparable successfully-relaxed GFN2-xTB CSD structures (Supplementary Fig. 10 and Supplementary Table 8). Accounting for unsuccessful generations that do not result directly from *Architector* failures (Supplementary Note 9), only 57 structures were unable to be handled by *Architector* construction, corresponding to a 99.1% success rate in generation. Comparable statistics on organometallic structure generation for the *CORINA* program⁵⁴ was 51.6% successful generation, while the *COSMOS* program reported comparable 94.6%, the only other programs benchmarked in a similar manner²⁷. Importantly, this prior benchmark organometallic set²⁷ did not include any structures with CN > 8, where actinides and lanthanides constitute the majority of observed complexes (Supplementary Fig. 8). Thus, *Architector* is directly competitive with state-of-the-art comparable programs even on complex actinide and lanthanide structures.

Focusing further on the f-block structures generated, we need reliable and descriptive methods for describing how close the generated structures are to the CSD reference structures. A tricky aspect of comparing different organometallic structures in terms of energy and structures is that variations in soft trailing organic components of the molecules can obscure a direct RMSD comparison between two organometallics, leading to identical graphs and visually/chemically similar structures being artificially labeled as distinct⁷. To directly compare the structures, we leverage the kmRMSD method specifically in the region proximal to the metal center within two graph hops of the metal center on the molecular graph (Fig. 3a). The kmRMSD up to depth two from the metal center was selected (c-kmRMSD), as this depth shows a stronger correlation with energy differences (Supplementary Fig. 11).

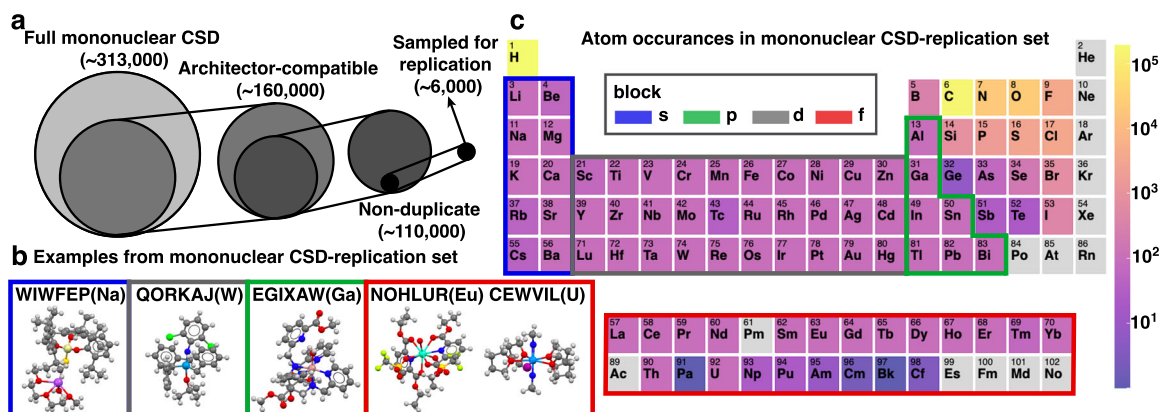


Fig. 2 | Cross-periodic table replication dataset construction. **a** Size of different subsets of the CSD resulting in the sample set for replication. **b** Examples structures of sampled replication set with Cambridge structural database (CSD) refcode(-metal) indicated. **c** Atom distributions present across all structures sampled from

the CSD divided by blocks. Inset images are molecules colored by atom type with Na:light purple, F/Cl:green, I:dark purple, N:blue, C:gray, Si:light yellow, O:red, S:yellow, H:white, Ga:brown, W/Eu/U:light blue.

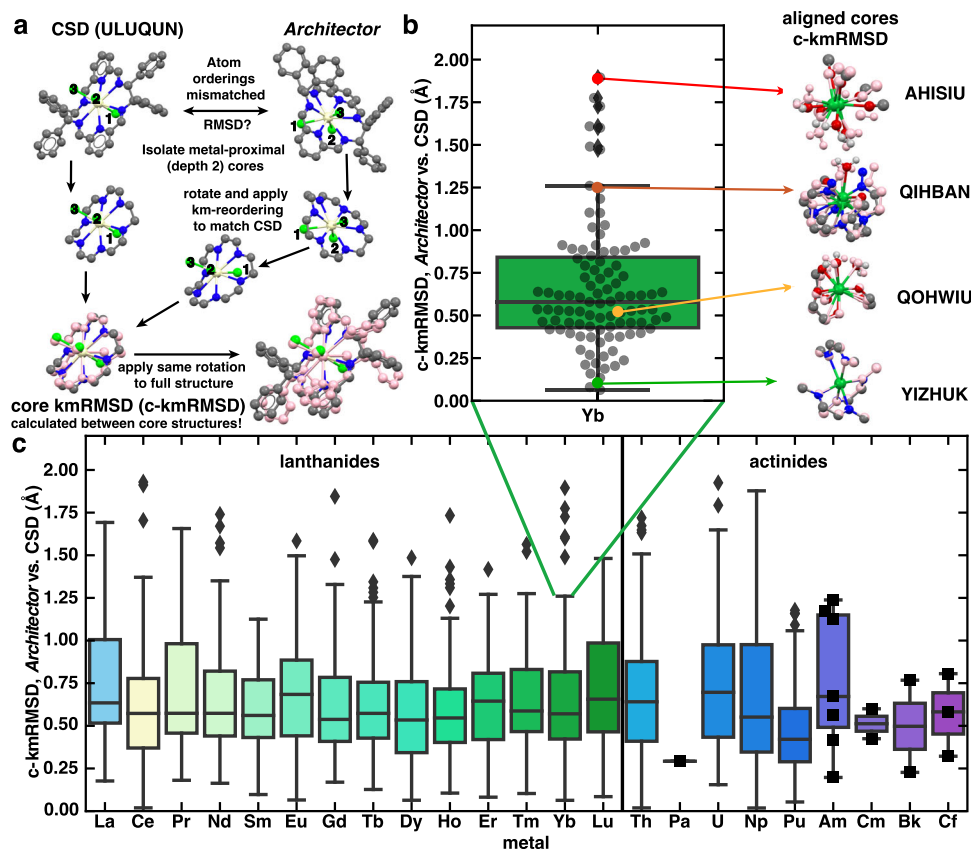


Fig. 3 | Replication of f-block CSD structures. **a** Illustration of core Kuhn–Munkers root-mean-squared distance (c-kmRMSD) alignment procedure **b** Single boxplot distribution of minimum-c-kmRMSD Architector-generated structures vs. CSD c-kmRMSD values for Ytterbium complexes. C-kmRMSD values are calculated only for the atoms within depth 2 neighbors of the metal center. Insets highlight closely mapped, average, and less-ideally mapped generated cores. **c** Boxplot distributions of c-kmRMSD of all minimum-c-kmRMSD Architector-generated lanthanide and actinide complexes. All insets the CSD structures are colored by atom type with

Yb:green, Cl:light green, N:blue, C:gray, O:red, Ce:yellow, H:white. Overlaid minimum-c-kmRMSD Architector-generated structures are colored pink to reveal differences in structure. Boxplots contain a colored region indicating the inner quartile range (IQR) between the 1st and 3rd quartiles, while whiskers indicate points within 1.5*(IQR) of either the 25 and 75 percentile points, and black diamonds indicate outliers beyond the 1.5*IQR cutoffs. Metals with fewer than 10 structures show all datapoints in black squares.

To illustrate how c-kmRMSD depicts differences in structures and how it trends with chemical intuition, we focus on Yb-generated compounds showing a wide spread in c-kmRMSD. For example, the Architector-generated Yb structure with c-kmRMSD near 2.00 Å has a metal-center symmetry with different ligand relative locations (Fig. 3b). Decreasing c-kmRMSD to near 1.25 Å reveals a non-distorted Yb structure with only one different relative ligand position than the CSD-derived structure. Looking at a structure near the mean c-kmRMSD (0.5 Å for Yb) reveals a close-to-perfect match in structure with only deviating hydrogen positions. Finally, looking at an example Yb structure with c-kmRMSD near 0.1 Å reveals virtually identical core and full structures. Broadly, this agrees with previous reports of RMSD similarity where RMSD < 1 Å can be regarded as identical conformers²⁷. Vitrally, higher c-kmRMSD often does not imply chemically unreasonable structures, merely differing ligand symmetries. Utilizing c-kmRMSD thus allows us to compare metal-centered structural comparison between generated and CSD-reference structures.

With c-kmRMSD, we primarily focused on the lanthanide/actinide f-block targets looking to generated conformers closest to the CSD equivalents (Fig. 3c). We immediately note that across the lanthanides and actinides, minimum c-kmRMSD values average near 0.5 Å indicating Architector typically generates at least one structure with minimal deviations between the generated structures and CSD structures. Lanthanides and actinides missing from Fig. 3c are not found at all in the CSD (Supplementary Fig. 8). We see that though actinides were

simulated in GFN2-xTB by replacing them with their f-electron equivalent lanthanides, the average c-kmRMSDs of actinides are comparable to those across the lanthanides. Additionally, only 35 of 1,645 f-block structures fall in the >1.5 Å c-kmRMSD region where the matching ligand symmetry to the CSD was not generated via Architector. We further note that across the whole periodic table there is similar average and worst-case c-kmRMSD values between Architector-generated and CSD-derived structures (Supplementary Fig. 12). Overall, these results highlight a great degree of agreement between the most similar Architector-generated complexes to CSD-derived structures they are replicating.

Beyond structural comparisons, energetic comparisons via GFN2-xTB can also be made between the Architector-generated structures and those from the CSD. With trailing-ligand conformations potentially playing a larger role in determining energetics than metal-proximal structure, we expect deviations in energetics between minimum-energy Architector-generated structures vs. CSD structures. We find instead that the mean absolute xTB energy difference on the minimum-energy Architector structures and the CSD structures is 12.1 kcal/mol (Fig. 4) with 96.1% of structure energies within 50 kcal/mol of the CSD structure. Additionally, the average difference between the CSD and Architector structures is -4.6 kcal/mol, indicating the Architector structures are, on average, only slightly higher in GFN2-xTB energy than the CSD structures. The trends and histogram of lanthanide energetics is also closely mapped by s-, p-, and d- block

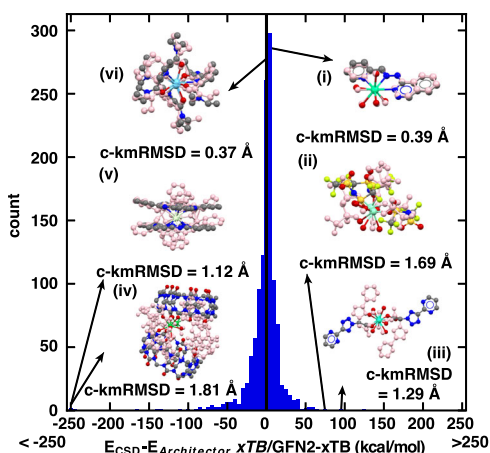


Fig. 4 | Lanthanide CSD vs. *Architector* GFN2-*xTB* energetics. Histogram of the difference in energies of the GFN2-*xTB*-relaxed CSD structure and the lowest-energy *Architector* conformers on lanthanide complex energetics. (i)–(vi) highlight examples of structures falling into extreme categories. (i) and (vi) represent *Architector* structures nearly isoenergetic with the CSD structure while (ii) and (iii) are > 50 kcal/mol more stable than the CSD structure and (iv) and (v) are > 250 kcal/mol less stable than the CSD structure. All insets the CSD structures are colored by atom type with Lns:green/light blue, F:light green, N:blue, C:gray, O:red, S:yellow. Overlaid minimum-*xTB*-energy *Architector*-generated structures are colored pink to reveal differences in structure with hydrogens removed for clarity. Associated c-kmRMSD values on the insets provide a measure of metal-centered similarity between the structures.

Architector-generated complexes as well (Supplementary Fig. 13). Overall, the *xTB* energetics of minimum-energy *Architector* structures are comparable with CSD structure energetics.

Instead of overall energetic trends, focusing on specific cases that both match and deviate in energy reveals potential sources of larger energetic discrepancies between *Architector* and the CSD. *Architector* structures much lower in energy (>50 kcal/mol lower) than the CSD structures seem to have bulkier ligands for which *Architector* identifies different ligand orientations that are lower in energy (Fig. 4ii/iii). These structures show c-kmRMSD values >1 Å indicating the core geometries are nearly identical, but ligand conformations differ, indicating crystal packing likely influenced the CSD conformer structure⁵⁵. Alternatively, structures for which the lowest-energy *Architector* structure was much higher in energy (>250 kcal/mol higher) than the CSD structures seem to be cases where *Architector*-generated a distorted ligand conformation different from that observed in the CSD (Fig. 4iv/v), albeit close core structures. However, only 33 out of 344 lanthanide structures with more than 100 atoms show >50 kcal/mol higher energy than the CSD structure. The structures much higher in energy than the CSD tend to have larger ligands, which have more rotational/structural degrees of freedom.

The *CREST* utility represents a powerful conformer sampling utility also applicable to lanthanide complexes targeted to sample local rotational/structural degrees of freedom. Comparison of *CREST* sampling performed directly on 19 CSD structures to *Architector*-generated structures revealed *CREST* generating structures lower in energy than the CSD structure in 83% of the cases with energy distributions on the order of 10 kcal/mol (Supplementary Note II and Supplementary Fig. 14). Since *Architector* represents a structure search over more diverse configurations, it produced structures lower in energy than the CSD in 84% of cases, with energy distributions on the order of 20 kcal/mol. The combination of these utilities, with *CREST* sampling performed after *Architector* generation is already implemented in *Architector* and is very likely to produce structures at or lower in energy than the CSD for a given molecule, though it

typically adds between 10–15 min of run time per complex to the generation procedure. We note that even without any additional conformational sampling via *CREST*, *Architector* identifies a majority (55.1%) of conformers for f-block lanthanides within typical *xTB* accuracy (2 kcal/mol)³¹ or lower in energy than the CSD structure. Additionally, although outliers in energetics exist, 64.5% of lowest-energy lanthanide *Architector* structures are within 10 kcal/mol (nearly isoenergetic and isostructural) with the CSD structures (Fig. 4i/vi). Finally, *CREST* requires a user-estimated 3D input structure for a molecule to perform conformer sampling, while *Architector* only requires 2D graph information about the molecule as input and thus can be easily applied to any metal-ligand combinations.

Larger *Architector* conformer space

Besides lowest-energy conformers, *Architector* can generate ranked-order conformers by GFN2-*xTB* energy. Such higher-energy conformers can be vital for understanding and accelerating f-block complex design. From a randomly-selected set of 765 complexes from across the periodic table we isolated 3,039 conformers generated by *Architector* to test with DFT relaxation methods (Methods Section and Supplementary Fig. 15). We note that since GFN2-*xTB* does not support atoms with $Z > 88$, an approximation for the actinides is needed for comparison of GFN2-*xTB* conformer energetics and structures. In *Architector* and for comparison to the CSD, actinides are approximated by swapping the actinide with the same-group lanthanides.

Crucially, in this work we are not intending to focus on accuracy of post-*Architector* methods or even the underlying *xTB* methods. From the electronic structure standpoint, we instead are focused on whether the *Architector*-generated structures are able to be evaluated by general electronic structure methods, and whether, from a given method these structures are energetically and geometrically sensible compared to experimentally-derived structures. For more detailed studies of specific regions of the periodic table any available method can be utilized to evaluate the configurations following *Architector* generation ranging from force fields (UFF) to complete active space self-consistent field (CASSCF). Users can readily customize both *Architector* internal methods (e.g., ASE Calculator (ADF/Gaussian/VASP), functional, and/or basis set) used and final evaluation methods to their own purposes. Given this broader goal and a desire to minimize computational expense, we have chosen to report PBE⁵⁶ + D^{57,58}/DZP methods⁵⁹ as the baseline *xTB* vs. DFT comparison (Fig. 5).

Over *Architector* structures, we were able to perform single point calculations on as many as 97% of all conformers with DFT (including 94% of lanthanide and 99% of actinide conformers attempted (Supplementary Fig. 15). Looking instead at distinct complexes, we were able to perform at least one successful DFT relaxation from 99% of lanthanide complexes and 100% of actinide complexes. Lanthanide complexes are known to be particularly difficult to converge in DFT^{60–62}, so a lower rate of success is to be expected and is consistent with convergence on CSD structures themselves (Supplementary Fig. 15). Thus, for most f-block complexes, and especially actinide complexes (even considering the approximation of swapping for lanthanides in GFN-*xTB* evaluation), *Architector* is generating structures amenable to DFT minimization.

Beyond the minimum-energy PES structures, we benchmarked *Architector* and the underlying GFN2-*xTB* ability to rank conformers in terms of energy and structure (Fig. 5). Over the full set of complexes and conformers evaluated with DFT (3039 conformers), we see that often the closest c-kmRMSD structure (vs. CSD) is not the lowest-energy for either DFT or GFN2-*xTB* (Fig. 5a). This again supports previous observations that minimized gas and liquid phase complex conformations do not necessarily correspond to those observed in single-crystal structures²⁸. These observations suggest that the distribution of conformers generated by *Architector* is key for capturing configurations important for different theoretical methods.

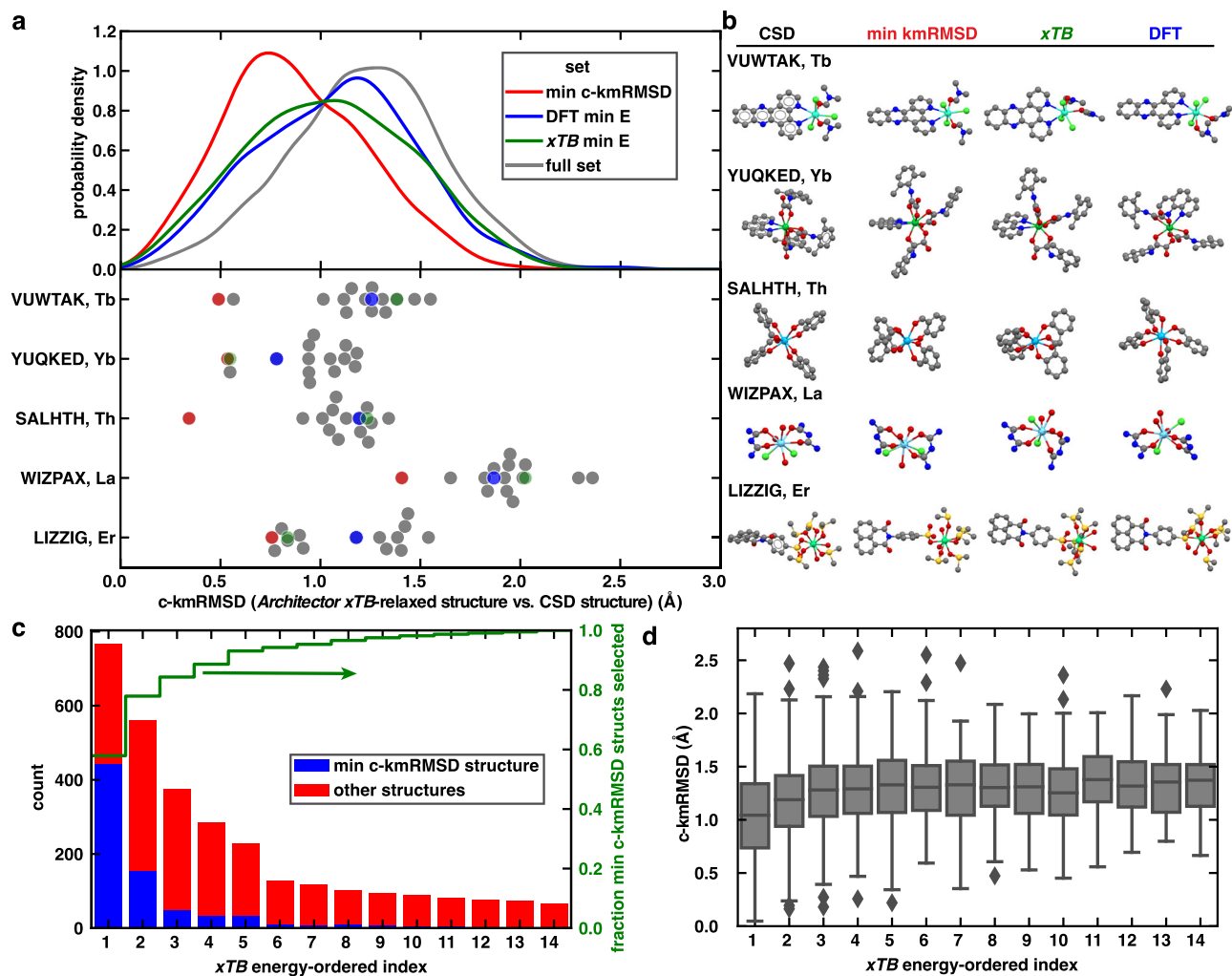


Fig. 5 | Conformer distributions and examples. **a** Kernel density estimate (KDE) plot of distributions of *Architector*-generated un-relaxed conformer core kmRMSD vs. CSD structures over the full set of 3039 structures from 765 different complexes (upper). (lower) Specific examples of structures where more than 13 conformers were relaxed by both DFT (PBE + D/DZP) and *xTB* (GFN2-*xTB*) and the distribution of their core kmRMSD values. **b** Renderings of the conformations corresponding to the structure directly from the CSD under the CSD refcode, the closest *Architector*-

generated un-relaxed initial conformer, the minimum-energy *xTB*-relaxed conformer, and the minimum-energy DFT-relaxed conformer. Calculations for actinides in *xTB* are reported with a lanthanide used as proxy. All renderings are colored by atom type with Lns:green/light blue, Cl:light green, N:blue, C:gray, O:red, S:yellow. **c** *xTB* energy-ordered index vs. fraction of minimum c-kmRMSD conformers and **d** *xTB* energy-ordered index vs. c-kmRMSD for all conformers evaluated with both *xTB*/DFT.

Looking at specific examples where more than thirteen conformers were able to be relaxed also provides insight into factors dictating minimum-energy structures for each method (Fig. 5b). Here, we directly see that *Architector* is typically generating at least one conformer quite close in structure to the CSD, but also forming a wide distribution of conformers in terms of metal-core local geometry (Fig. 5a). Both GFN2-*xTB* and DFT favor structures with two key differences from the CSD. First, for structures DFT/GFN2-*xTB* VUWTAK, SALHTH, LIZZIG, and WIZPAX, different ligand mappings and metal-center symmetries are observed from the CSD structure (Fig. 5b). This type of deviation potentially agrees with previous catalytic work where the catalytically active and lowest-energy structures had different relative ligand symmetries⁶³. We note that accessing these types of conformers directly from an XRD initial structure is difficult, if not impossible³⁴. The second type of deviation is primarily observed in the YUQKED structures, where the metal-ligand symmetry is nearly the same as the experimental structure, but trailing-ligand orientation differs resulting in a lower-energy conformer. Additionally, we note that oxidation states for lanthanides and actinides differing from the +3 state were also sampled and *Architector*-produced structures for

these compounds were comparable to experimental structures (e.g., SALHTH in Fig. 5a/b). Since nearly all the multi-symmetry (more than 13 conformers) complexes are from the f-block, ligand symmetry effects appear especially important when considering actinide and lanthanide complexes as enabled in *Architector*.

Another key question relevant to structural generation is how many conformers are needed to produce structures closest to experimental structures. Here, we note that *xTB* energy-index ordering tends to trend strongly with minimum c-kmRMSD structures with over 93.3% of minimum c-kmRMSD structures being found in the top 5 lowest-*xTB*-energy conformers (Fig. 5c). Additionally, we directly observe that *xTB* energy-ordered index also trends directly with c-kmRMSD with lower indices tending to have lower c-kmRMSD values (Fig. 5d). Beyond closeness to experiments, *xTB*-ordered index also appears to trend with DFT-evaluated relative energy (Supplementary Fig. 16). Together, these observations indicate that the combination of *Architector*/*xTB* tends to select for structures that are close to experiments but potentially contain some key differences in ligand symmetry, resulting in the sampling of lower-energy structures.

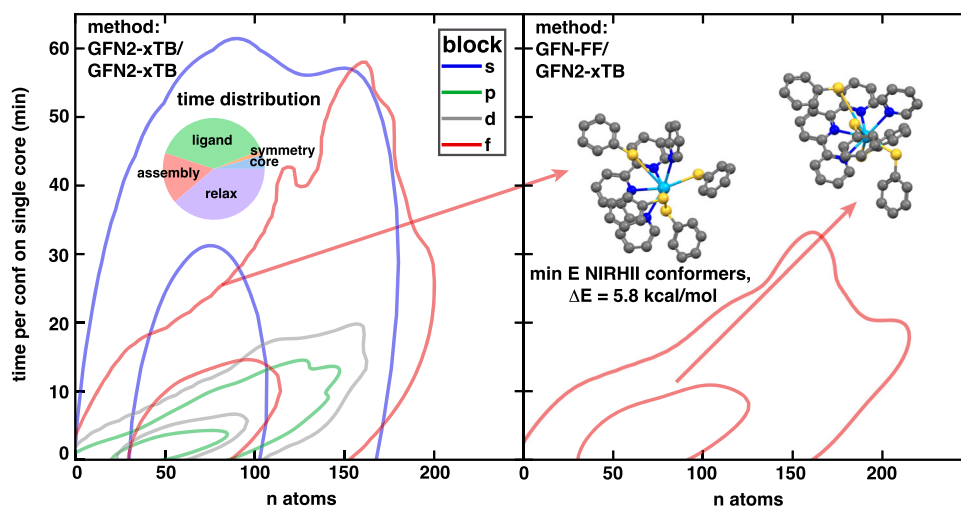


Fig. 6 | Timings and Performance of *Architector*. 2D Kernel Density Estimates (KDE) of the time per configuration on a single conformer on a single core vs. number of atoms (n atoms) by each block. The smaller circles represent density estimates of 50% of the data, while the outer circles represent 95% density estimates. The inset on the left indicates the average distribution of the times for each subtask of *Architector* where “core” refers to core and core-ligand mapping determination time, “symmetry” to finding valid sets of core-ligand bindings, “ligand” to

ligand conformer generation time, “assembly” to time placing and evaluating ligand conformers, and “relax” to the final relaxation time with GFN2-xTB. The right plot is the same generations performed with GFN-FF for assembly rather than GFN2-xTB for all Ce, Nd, Eu, Ho, Th, and U clusters. The inset highlights the structural results for both methods are nearly identical and are colored by atom type with Th:light blue, N:blue, C:gray, S:yellow.

High-throughput capabilities

Assessing the properties of yet-unsynthesized complexes requires rapid screening of large chemical spaces and numbers of conformers. From the set replicating the CSD, we report an average generation time of 6.7 min per valid conformer returned on a single core (Fig. 6 and Supplementary Fig. 17). For structures containing fewer than 100 atoms (~80% of structures), an average of 3.3 min per valid conformer on a single core is achieved. We note that the assembly and relaxation steps of the final complexes dominate the time spent generating complexes. Instead of using GFN2-xTB for both steps, a user can request the force field, e.g., GFN-FF⁶⁴ for the assembly and GFN2-xTB for relaxation, as leveraged in *CREST*³³. Over the same set of f-block complexes, simply swapping GFN2-xTB for GFN-FF during assembly resulted in a reduction in generation time by a factor of 2 with minimal tradeoff in the quality of the results (Fig. 6).

Architector can also be readily run in an “embarrassingly” parallel fashion over complexes, which enabled over 99% of the 6154 structural generations relaxed at the GFN2-xTB level to complete within 12 h on ~500 cores. Given the rapidity of other methods of 2D to 3D-generation such as *CORINA*, which reports around 0.009 min per conformer on a single core over *Pubchem*, this may still seem comparatively slow^{54,65}. However, given the complexity of the targeted systems (f-block) along with accounting for spin, charge, and energetics of multiple conformers (not performed with any other package), *Architector* is a very effective tool for high-throughput 2D to 3D generation for challenging complex organometallics across the periodic table.

Discussion

Architector represents a transformative tool for 3D structure generation of mononuclear organometallic compounds, particularly f-block complexes. We have shown it capable of producing chemically valid 3D structures from 2D inputs, producing numerous GFN2-xTB-relaxed conformers for each given input. The energetics and introduced core kmRMSD comparison between the CSD and *Architector*-generated f-block complexes reveal quite similar structures. Due to conformational sampling and crystal field effects, in a majority of cases *Architector* was able to find comparable-energy gas-phase structures to those obtained from the CSD and evaluated by both DFT and GFN2-xTB. With many

chemically-sensible conformers generated for each structure and energetic rankings of non-minima energy conformers, *Architector* is ready for use in tasks such as force field and tight binding training for chemical systems where cheaper methods do not yet exist^{31,66}. Finally, *Architector* shows strong promise for high-throughput structure generation of mononuclear complexes across the whole periodic table, which can enable the data-driven discovery of complex chemical systems essential for energy, catalysis, separation that are on the horizon.

Although *Architector* is achieving over 97% success in structural generation across the periodic table, as with any 3D-generation routine, *Architector* relies on heuristics and thus is not guaranteed to find the lowest-energy conformer for a given molecule. Further developments in *Architector* are underway to expand its capabilities to include ligand type handling for structures with neighbors bound to the metal center and haptic ligands with CAs not involved in the haptic interaction, representing ~7% of ligands found in the CSD (Supplementary Table 6). In this first version of *Architector*, we focus on mononuclear metal compounds, while extensions to polynuclear structures by linking mononuclear centers are also under development.

Overall, the structural generation and conformer ranking ability of *Architector* is remarkably broad. With conformers readily evaluated with any electronic structure method, the screening of metal spin, metal charge, ligand coordinating geometries, and ligand packing become available, enabling accurate evaluation of properties such as ligand binding energies, spectroscopic properties, and potentially catalytic properties of organometallics. Although primarily developed for higher-coordinated lanthanide and actinide chemistries, *Architector* has proven applicable across the periodic table, dramatically expanding its potential usage. Additionally, though benchmarked against known crystal structures, there is no limitation of *Architector* to known chemical structures as it is readily applied to any metal-ligand complex. Thus, *Architector* represents a vital step towards data-driven discovery in f-block chemistries and beyond.

Methods

For CSD replication calculations, up to 20 different core symmetries per core-ligand mapping were requested, with the top 10 lowest-energy conformers relaxed by *xTB*/GFN2-xTB. All assembly steps were

performed with *xTB*/GFN2-xTB as well except where otherwise indicated. Single point GFN2-xTB energetic evaluations with the methanol ($\epsilon = 33$)⁶⁷ solvent revealed shifts in conformer energetics and energetic rankings (Supplementary Figs. 18 and 19). Shifts are expected given more polar conformers would be more stabilized by the presence of solvents and are suggested for use in cases where solvents are known. In *Architector*, adding solvents supported by *xTB* during complex construction and conformer ranking is facilitated by adding a single “solvent” keyword. For this work, since the structures evaluated are derived from crystals made in different solvents (e.g., $\epsilon = 2$ for hexane to $\epsilon = 80$ for water)⁶⁷, simulating in the gas-phase serves as a common reference. Ligand types were assigned for replication by choosing the ligand type classified from the 3D CSD structure (Supplementary Table 2). Duplicates with full kmRMSD values lower than 0.5 Å were removed. *Architector* relies heavily on several open-source packages including *ASE*⁵⁰, *scipy*⁶⁸, and *xTB*⁶⁶. Internal visualization is supported using the *py3Dmol*⁶⁹ plugin for jupyter notebooks. All molecular renderings were made using the *Mercury*⁷⁰ package from the CCDC.

All DFT relaxations were performed with the *ADF* 2022.101⁷¹ using the PBE exchange-correlation density functional⁵⁶ managed by the open-source *pyiron* package⁷². For structures with all elements $Z < 86$, D4 dispersion corrections (+D)⁵⁷ were added, while for structures with all elements $Z < 95$, D3 corrections⁵⁸ were added to facilitate more similar comparison between DFT and *xTB* methods. A smaller double-zeta basis set with polarization functions (DZP)⁵⁹ was used to accelerate DFT calculations and encourage convergence, while ZORA scalar relativistic effects⁷³ were employed to capture relativistic effects typically present in f-block elements. Single point comparisons revealed minimal impact of basis set and functional selection on conformer ranking energetics (Supplementary Note 12 and Fig. 20). We evaluated spin contamination on the PBE + D/DZP DFT calculations and found only 7 of 3039 structures contained $\langle S^2 \rangle$ deviations greater than 0.5 (Supplementary Fig. 21).

Data availability

All data generated in this study have been deposited in the Zenodo database under accession code <https://doi.org/10.5281/zenodo.7764697>. The data contained in the figures in this study are provided in the Source Data file. Source data are provided with this paper.

Code availability

The *Architector* code is available on [github](https://github.com) at and is also able to be installed via the Anaconda (conda install -c conda-forge *architector*) or pip (pip install *architector*) package managers. A permanent version associated with this manuscript can be found at <https://doi.org/10.5281/zenodo.7754279>.

References

- Pollice, R. et al. Data-driven strategies for accelerated materials design. *Acc. Chem. Res.* **54**, 849–860 (2021).
- Martin, J. H. et al. 3D printing of high-strength aluminium alloys. *Nature* **549**, 365–369 (2017).
- Singh, A. K., Montoya, J. H., Gregoire, J. M. & Persson, K. A. Robust and synthesizable photocatalysts for CO₂ reduction: a data-driven materials discovery. *Nat. Commun.* **10**, 443 (2019).
- Schaduangrat, N. et al. Towards reproducible computational drug discovery. *J. Cheminform.* **12**, 9 (2020).
- Gao, W., Raghavan, P. & Coley, C. W. Autonomous platforms for data-driven organic synthesis. *Nat. Commun.* **13**, 1075 (2022).
- Ioannidis, E. I., Gani, T. Z. H. & Kulik, H. J. molSimplify: a toolkit for automating discovery in inorganic chemistry. *J. Comput. Chem.* **37**, 2106–2117 (2016).
- Sobez, J.-G. & Reiher, M. Molassembler: molecular graph construction, modification, and conformer generation for inorganic and organic molecules. *J. Chem. Inf. Model.* **60**, 3884–3900 (2020).
- Foscato, M., Venkatraman, V. & Jensen, V. R. DENOPTIM: software for computational de novo design of organic and inorganic molecules. *J. Chem. Inf. Model.* **59**, 4077–4082 (2019).
- Bennett, W. E. Computation of the number of isomers and their structures in coordination compounds. *Inorg. Chem.* **8**, 1325–1328 (1969).
- Riniker, S. & Landrum, G. A. Better informed distance geometry: using what we know to improve conformation generation. *J. Chem. Inf. Model.* **55**, 2562–2574 (2015).
- Munguba, G. H. L. et al. The complex build algorithm to set up starting structures of lanthanoid complexes with stereochemical control for molecular modeling. *Sci Rep* **11**, 21493 (2021).
- Paiva, A. P. & Malik, P. Recent advances on the chemistry of solvent extraction applied to the reprocessing of spent nuclear fuels and radioactive wastes. *J. Radioanal. Nucl. Chem.* **261**, 485–496 (2004).
- Veliscek-Carolan, J. Separation of actinides from spent nuclear fuel: a review. *J. Hazard. Mater.* **318**, 266–281 (2016).
- Hidayah, N. N. & Abidin, S. Z. The evolution of mineral processing in extraction of rare earth elements using liquid-liquid extraction: a review. *Miner. Eng.* **121**, 146–157 (2018).
- Balaram, V. Rare earth elements: a review of applications, occurrence, exploration, analysis, recycling, and environmental impact. *Geosci. Front.* **10**, 1285–1303 (2019).
- Yaroshevsky, A. A. Abundances of chemical elements in the Earth’s crust. *Geochem. Int.* **44**, 48–55 (2006).
- Burns, C. J. & Eisen, M. S. *The Chemistry of the Actinide and Transactinide Elements*. 2799–2910 (Springer).
- Pepper, M. & Bursten, B. E. The electronic structure of actinide-containing molecules: a challenge to applied quantum chemistry. *Chem. Rev.* **91**, 719–741 (1991).
- Su, J. et al. Energy-degeneracy-driven covalency in actinide bonding. *J. Am. Chem. Soc.* **140**, 17977–17984 (2018).
- Mason, J. L., Harb, H., Topolski, J. E., Hratchian, H. P. & Jarrold, C. C. Exceptionally complex electronic structures of lanthanide oxides and small molecules. *Acc. Chem. Res.* **52**, 3265–3273 (2019).
- Cantu, D. C. Chapter Seven - Predicting lanthanide coordination structures in solution with molecular simulation. In *Rare-Earth Element Biochemistry: Characterization and Applications of Lanthanide-Binding Biomolecules*, Vol. 651 (ed. Cotruvo, J. A.) 193–233 (Academic Press, 2021).
- Morgenstern, A. et al. Computer-assisted design of macrocyclic chelators for actinium-225 radiotherapeutics. *Inorg. Chem.* **60**, 623–632 (2021).
- Arumugam, K. & Burton, N. A. Density functional theory (DFT) calculations of VI/V reduction potentials of uranyl coordination complexes in non-aqueous solutions. *Phys. Chem. Chem. Phys.* **21**, 3227–3241 (2019).
- Janet, J. P., Duan, C., Nandy, A., Liu, F. & Kulik, H. J. Navigating transition-metal chemical space: artificial intelligence for first-principles design. *Acc. Chem. Res.* **54**, 532–545 (2021).
- Balcells, D. & Skjelstad, B. B. tmQM dataset—quantum geometries and properties of 86k transition metal complexes. *J. Chem. Inf. Model.* **60**, 6135–6146 (2020).
- Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge structural database. *Acta Crystallogr. Sect. B Struct. Sci., Cryst. Eng. Mater.* **72**, 171–179 (2016).
- Andronico, A., Randall, A., Benz, R. W. & Baldi, P. Data-driven high-throughput prediction of the 3-d structure of small molecules: review and progress. *J. Chem. Inf. Model.* **51**, 760–776 (2011).
- Mitzel, N. W. & Rankin, D. W. H. SARACEN-molecular structures from theory and experiment: the best of both worlds. *Dalton Trans.* **2003**, 3650–3662 (2003).

29. Grimme, S. et al. Fully automated quantum-chemistry-based computation of spin-spin-coupled nuclear magnetic resonance spectra. *Angew. Chem. Int. Ed.* **56**, 14763–14769 (2017).
30. Pracht, P., Wilcken, R., Udvarhelyi, A., Rodde, S. & Grimme, S. High accuracy quantum-chemistry-based calculation and blind prediction of macroscopic pKa values in the context of the SAMPL6 challenge. *J. Comput.-Aided Mol. Des.* **32**, 1139–1149 (2018).
31. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).
32. Bursch, M., Neugebauer, H. & Grimme, S. Structure optimisation of large transition-metal complexes with extended tight-binding methods. *Angew. Chem. Int. Ed.* **58**, 11078–11087 (2019).
33. Pracht, P., Bohle, F. & Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **22**, 7169–7192 (2020).
34. Bursch, M., Hansen, A., Pracht, P., Kohn, J. T. & Grimme, S. Theoretical study on conformational energies of transition metal complexes. *Phys. Chem. Chem. Phys.* **23**, 287–299 (2021).
35. Dam, H. H., Reinhoudt, D. N. & Verboom, W. Multicoordinate ligands for actinide/lanthanide separations. *Chem. Soc. Rev.* **36**, 367–377 (2007).
36. Spezia, R., Migliorati, V. & D’Angelo, P. On the development of polarizable and Lennard-Jones force fields to study hydration structure and dynamics of actinide(III) ions based on effective ionic radii. *J. Chem. Phys.* **147**, 161707 (2017).
37. Carlson, R. K., Cawkwell, M. J., Batista, E. R. & Yang, P. Tight-binding modeling of uranium in an aqueous environment. *J. Chem. Theory Comput.* **16**, 3073–3083 (2020).
38. Nguyen, M.-T., Rousseau, R., Paviet, P. D. & Glezakou, V.-A. Actinide molten salts: a machine-learning potential molecular dynamics study. *ACS Appl. Mater. Interfaces* **13**, 53398–53408 (2021).
39. Lubbers, N., Smith, J. S. & Barros, K. Hierarchical modeling of molecular energies using a deep neural network. *J. Chem. Phys.* **148**, 241715 (2018).
40. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **28**, 31–36 (1988).
41. Thomassen, C. The graph genus problem is NP-complete. *J. Algorithms* **10**, 568–576 (1989).
42. O’Boyle, N. M. et al. Open Babel: an open chemical toolbox. *J. Cheminformatics* **3**, 33 (2011).
43. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17**, 490–519 (1996).
44. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).
45. Blaney, J. M. & Dixon, J. S. in *Reviews in Computational Chemistry* 299–335 (VCH, 1994).
46. Crippen, G. & Havel, T. F. *Distance Geometry and Molecular Conformation* (Research Studies Press; Wiley, 1988).
47. Pyykkö, P. Additive covalent radii for single-, double-, and triple-bonded molecules and tetrahedrally bonded crystals: a summary. *J. Phys. Chem. A* **119**, 2326–2337 (2015).
48. Alvarez, S. A cartography of the van der Waals territories. *Dalton Trans.* **42**, 8617 (2013).
49. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* **32**, 922–923 (1976).
50. Hjorth Larsen, A. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.: Condens. Matter* **29**, 273002 (2017).
51. Temelso, B., Mabey, J. M., Kubota, T., Appiah-Padi, N. & Shields, G. C. ArbAlign: a tool for optimal alignment of arbitrarily ordered isomers using the Kuhn–Munkres algorithm. *J. Chem. Inf. Modeling* **57**, 1045–1054 (2017).
52. Taylor, M. G. et al. Seeing is believing: experimental spin states from machine learning model structure predictions. *J. Phys. Chem. A* **124**, 3286–3299 (2020).
53. Mentel, L. mendeleev—a Python resource for properties of chemical elements, ions and isotopes v. 0.12.1 (2014).
54. Molecular Networks GmbH, Germany & Altamira, LLC, USA. *CORINA Classic v. 4.4.0* (Altamira LLC & Molecular Networks GmbH, 2021).
55. Martín, A. & Orpen, A. G. Structural systematics. 6.1 apparent flexibility of metal complexes in Crystals. *J. Am. Chem. Soc.* **118**, 1464–1470 (1996).
56. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
57. Caldeweyher, E. et al. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **150**, 154122 (2019).
58. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
59. Van Lenthe, E. & Baerends, E. J. Optimized Slater-type basis sets for the elements 1–118. *J. Comput. Chem.* **24**, 1142–1156 (2003).
60. Calvello, S., Piccardo, M., Rao, S. V. & Soncini, A. CERES: an ab initio code dedicated to the calculation of the electronic structure and magnetic properties of lanthanide complexes. *J. Comput. Chem.* **39**, 328–337 (2018).
61. Martínez-Flores, C., Bolívar-Pineda, L. M. & Basiuk, V. A. Lanthanide bisphthalocyanine single-molecule magnets: A DFT survey of their geometries and electronic properties from lanthanum to lutetium. *Mater. Chem. Phys.* **287**, 126271 (2022).
62. Basiuk, V. A., Prezhdo, O. V. & Basiuk, E. V. Thermal smearing in DFT calculations: How small is really small? A case of La and Lu atoms adsorbed on graphene. *Mater. Today Commun.* **25**, 101595 (2020).
63. Yuan, M., Song, Z., Badir, S. O., Molander, G. A. & Gutierrez, O. On the nature of C(sp³)–C(sp²) bond formation in nickel-catalyzed tertiary radical cross-couplings: a case study of ni/photoredox catalytic cross-coupling of alkyl radicals and aryl halides. *J. Am. Chem. Soc.* **142**, 7225–7234 (2020).
64. Spicher, S. & Grimme, S. Robust atomistic modeling of materials, organometallic, and biochemical systems. *Angew. Chem. Int. Ed.* **59**, 15665–15673 (2020).
65. Kim, S. et al. PubChem substance and compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2016).
66. Bannwarth, C. et al. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* **11** <https://doi.org/10.1002/wcms.1493>, (2021)
67. Rumble, J. R. (ed.) in *CRC Handbook of Chemistry and Physics, 103rd Edition (Internet Version 2022)* (CRC Press/Taylor & Francis, Boca Raton, FL, 2022).
68. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

69. Rego, N. & Koes, D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* **31**, 1322–1324 (2015).
70. Bruno, I. J. et al. New software for searching the Cambridge structural database and visualizing crystal structures. *Acta Crystallogr. Sect. B Struct. Sci.* **58**, 389–397 (2002).
71. Te Velde, G. et al. Chemistry with ADF. *J. Comput. Chem.* **22**, 931–967 (2001).
72. Janssen, J. et al. pyiron: An integrated development environment for computational materials science. *Comput. Mater. Sci.* **163**, 24–36 (2019).
73. Van Lenthe, E., Ehlers, A. & Baerends, E.-J. Geometry optimizations in the zero order regular approximation for relativistic effects. *J. Chem. Phys.* **110**, 8943–8953 (1999).

Acknowledgements

We acknowledge the support by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Separation Science Program under contract number KC0302020 (M.G.T., D.J.B) and Heavy Element Chemistry Program under contract KC0302031 at Los Alamos National Laboratory (LANL) (J.J., E.R.B., D.P., P.Y.). LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (contract no. 89233218CNA000001). We acknowledge the LANL's Director's Postdoc Fellowship (LANL-LDRD, 20210966PRD4) (M. G. T) and LANL's Center for Nonlinear Studies (CNLS) Postdoc Fellowship (D.J.B.). The computation works were supported by LANL Institution Computing. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAPO022265.

Author contributions

M.G.T. and D.J.B. developed the Architector code and wrote the initial drafts of the manuscript. J.J. aided in data collection and Architector codebase. D.P., E.R.B., and P.Y. conceived of the project. All authors aided in the writing and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-38169-2>.

Correspondence and requests for materials should be addressed to Enrique R. Batista, Danny Perez or Ping Yang.

Peer review information *Nature Communications* thanks Jan-Michael Mewes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023, corrected publication 2023