



From molecular generation to process-relevant performance: AI-driven identification of novel working fluids

Manuel Ruiz-Botella ^{a,1}, Tiago M. Eusébio ^{a,b,1}, Marta Sales-Pardo ^{a,c,1,*},
Roger Guimerà ^{a,c,d,1,*}, Sabrina Belén Rodríguez-Reartes ^{a,1}, Fèlix Llovell ^{a,1,*}

^a Department of Chemical Engineering, Universitat Rovira i Virgili, Campus Sescelades, Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain

^b Department of Chemical Engineering and Materials Science, IQS School of Engineering, Universitat Ramon Llull, Via Augusta 390, 08017, Barcelona, Spain

^c Center for Computational Science and Applied Mathematics (ComSCIAM), Universitat Rovira i Virgili, 43007 Tarragona, Catalonia, Spain

^d ICREA, 08010 Barcelona, Catalonia, Spain

ARTICLE INFO

Dataset link: <https://github.com/manurubo/cograph>, <https://github.com/manurubo/refrig-erant-screening-workflow>

Keywords:

Generative molecular design
Graph diffusion models
Working fluid discovery
Molecular dynamics simulation
Molecular-based equations of state
Thermal cycle applications

ABSTRACT

The search for working fluids that balance thermodynamic performance, safety, and environmental sustainability remains a pressing challenge for refrigeration and heat pump technologies. Here, we present a hybrid computational workflow combining deep generative learning with classical thermodynamic modelling to identify novel working fluid candidates. Using a graph-based diffusion model, we generate and screen molecules against a quantitative physicochemical profile derived from known working fluids. We characterize five selected candidates in depth through molecular dynamics simulations, the polar soft-SAFT EoS, COSMO-RS calculations, and two neural network predictors for flammability and global warming potential (GWP). We find that the predicted thermophysical and transport properties of the candidates fall within the range of established hydrofluorocarbons. Benchmarking against R-245fa and R-134a identifies a new compound (3,3,3-trifluoro-2-methylpropene) as the most promising candidate for high-temperature heat pump applications, and confirms another (1,1,1,2,2-pentafluoropropane) as a viable refrigerant for vapour compression refrigeration cycles. Our work establishes a scalable framework for the rational, data-driven discovery of next-generation working fluids.

1. Introduction

Working fluids are fundamental to the operation of modern thermal technologies. Refrigerators, heat pumps, and air-conditioning systems across automotive, domestic, commercial, and industrial sectors rely on these fluids, which undergo cyclic phase changes between liquid and vapour to absorb, transport, and release heat efficiently. Refrigeration itself stands among the major technological achievements of the past century, playing a critical role in the global economy, impacting food preservation, public health and energy systems. However, its global footprint is substantial, consuming 20% of the world's electricity and accounting for 7.5% of global CO₂ emissions, from which 30% are consequence of fluid leaks [1].

This growing energetic and environmental burden has shifted attention toward the choice of adequate working fluids whose development has been increasingly guided by regulatory and sustainability constraints rather than solely by performance factors [2–5]. Early vapor

compression refrigeration cycle (VCRC) systems relied on common volatile fluids that were often toxic and flammable [6]. These were replaced in the 1930s by chlorofluorocarbons (CFCs), which combined excellent thermodynamic performance, chemical stability and low toxicity. However, their severe ozone depletion potential (ODP) led to their phase-out under the Montreal Protocol [7]. Subsequent generations, including hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (HFCs), reduced or eliminated ozone depletion but introduced significant global warming concerns, ultimately leading to regulatory measures such as the Kigali Amendment [8]. More recently, unsaturated fluorinated compounds, such as hydrofluoroolefins (HFOs), have emerged as low global warming potential (GWP) alternatives with negligible ODP. However, they present trade-offs, including flammability, lower volumetric cooling capacity, and potential environmental concerns related to their degradation into persistent substances such as trifluoroacetic acid [9].

* Corresponding authors.

E-mail addresses: manuel.ruiz@urv.cat (M. Ruiz-Botella), tiago.mendonca@iqs.url.edu (T.M. Eusébio), marta.sales@urv.cat (M. Sales-Pardo), roger.guimera@urv.cat (R. Guimerà), sabrinabelen.rodriguez@urv.cat (S.B. Rodríguez-Reartes), felix.llovell@urv.cat (F. Llovell).

¹ These authors contributed equally to this work.

Despite significant advances, no existing working fluid fulfils all the required criteria in terms of performance, safety, and environmental sustainability. At the same time, increasingly strict climate policies are enforcing the phase-down of high-GWP working fluids [10,11], even though most existing cooling and heating systems still depend heavily on them. This makes the search for alternative working fluids that satisfy environmental and safety requirements, while also exhibiting favourable thermodynamic and transport properties, a pressing technological challenge [5]. However, identifying such fluids is far from straightforward, as the number of viable candidates appears to be inherently limited. In a comprehensive screening study, McLinden et al. assessed a database of more than 60 million single component fluids and found that only 27 possessed any realistic potential as low-GWP working fluids [12]. The magnitude of this challenge is further illustrated by the fact that, since 2015, only a handful of new substances (including R-1130(E), R-1336mzz(Z), R-1336mzz(E), R-1233zd(E), R-1224yd(Z), R-131I, and R-1132a) were incorporated into the ANSI/ASHRAE Standard 34.

Given this, current efforts focus either on the characterization of fluid mixtures and their properties [13,14] or on the discovery of entirely new molecules. For the latter, machine learning is a promising tool, as the extraordinarily large size of the chemical compound space, estimated to comprise around 10^{60} different molecules [15], turns traditional trial-and-error discovery approaches inefficient and impractical. Indeed, fields such as drug discovery [16] and advanced materials development [17–19] are increasingly adopting artificial intelligence methods capable of generating new molecules with desired features, opening new opportunities also for the systematic design of next-generation working fluids. However, even though generative models can propose new molecules that are in principle suitable candidates, their real potential can only be assessed from the accurate estimation of thermodynamic and transport properties for which robust and predictive methods are necessary.

Here, we propose a hybrid framework combining generative deep learning techniques with classical thermodynamic methods to find new working fluids that meet market performance requirements, which we illustrate in Fig. 1.

First, we establish a physicochemical property profile characteristic of known refrigerants by comparing them against a diverse, size-matched molecular database. Next, we generate new molecular candidates using CoCoGraph [20] and extract a detailed suite of physicochemical descriptors (Fig. 1, Step 1). Then, we use the established profile to systematically screen millions of candidates and filter them based on their multidimensional distance to the working fluid profile (Fig. 1, Step 2).

Finally, from the resulting list of top candidates, we select a subset of five promising compounds to assess in depth their application feasibility (Fig. 1, Step 3). We perform atomistic molecular dynamics (MD) simulations to predict vapour–liquid equilibrium densities, vapour pressure, surface tension, enthalpy of vaporization, diffusivity, and viscosity. The simulation results serve as the basis for a thermodynamic model within the soft-Statistical Associating Fluid Theory (soft-SAFT) equation of state (EoS) framework [21]. Additionally, we conduct calculations using the CONductor-like Screening MOdel for Real Solvents (COSMO-RS) [22] to obtain the molecular surface charge density distributions and predict transport properties of the candidates. We also obtain the COSMO-RS [22] sigma profiles, and use them as input to a neural network model [14] capable of predicting the flammability of the selected compounds. Ultimately, we benchmark the thermodynamic and transport properties, as well as the predicted flammability and GWP, of the new compounds against those of the most common working fluids to assess their suitability for use in VCRCs and high temperature heat pumps (HTHPs), including preliminary cycle calculations.

2. Background

2.1. Molecular generative models

Developing machine learning approaches capable of generating novel plausible molecules is challenging because the problem is discrete, combinatorial, and constrained by the intricate chemical rules that realistic molecules must fulfil. Thus, early generative models for molecules using variational autoencoders [23], generative adversarial networks [24], and graph neural networks [25] faced challenges related to scalability, computational efficiency, adherence to chemical constraints, and ability to generalize [26–28].

In recent years, advances in graph-based diffusion models [29,30] and related approaches such as flow-matching [31] have alleviated such difficulties. In graph-based diffusion models for molecule generation, a controlled noising process is used to distort a molecular graph (by adding atoms and/or bonds); these noising trajectories are then used to learn the reverse denoising process, which reconstructs molecules from random molecular graphs and can ultimately be used to generate new molecules. The family of graph-based diffusion models for molecule generation includes DiGress [32], CDGS [33], and CoCoGraph [20]. CoCoGraph in particular is, by construction, guaranteed to generate valid molecules, and is also the model that generates synthetic molecules with the highest diversity and physicochemical properties more closely resembling those of real molecules [20].

2.2. Computational thermodynamics methods

Molecular-based theoretical approaches, most notably MD simulations and equations of state, such as the statistical association fluid theory (SAFT) [21], and quantum-based theories such as conductor-like screening model for real solvents (COSMO-RS) [22], provide powerful strategies to address the challenge of estimating key thermophysical properties of putative working fluids. These computational tools offer time- and cost-effective alternatives to experimental characterization while delivering reliable and detailed insights into the thermodynamic behaviour of fluids. MD simulations have proven to be a solid tool for predicting both bulk and interfacial properties across a broad range of fluid systems, including pure working fluids and their mixtures [34]. While extensive efforts have been devoted to the development of specific refrigerant force fields [35,36], general and transferable force fields, such as OPLS-AA [37], have also been successfully used to model working fluids [38,39].

The soft-SAFT EoS [40] has consistently demonstrated strong predictive capability for phase equilibria across a wide range of fluid systems, including HFCs and HFOs used as working fluids [41,42]. Furthermore, the spectrum of properties that can be modelled with SAFT-type equations can be easily extended by coupling them with other theoretical approaches. For example, SAFT-type models can be extended to the calculation of interfacial properties by coupling them with density gradient theory (DGT) [43] or to the calculation of viscosity by integrating the free-volume theory (FVT) [44]. Finally, COSMO-RS can be used to gain insight on the molecular geometry, polarity and charge distribution areas of any molecule, as well as make predictions of a set of fluid properties [45]. Beyond that, molecular descriptors generated by COSMO-RS can easily be integrated into machine-learning models, enabling the prediction of thermophysical properties [13,46], and the flammability [14] of conventional working fluids used in refrigeration applications.

It is important to note that Group-Contribution (GC)-SAFT approaches, particularly SAFT- γ Mie [47], which has demonstrated high accuracy across a broad range of systems, can be used as an alternative path to provide an accurate and computationally efficient alternative for predicting the thermodynamic properties of candidate working fluids without requiring prior molecular simulations. However, their application is restricted to molecules composed of functional groups

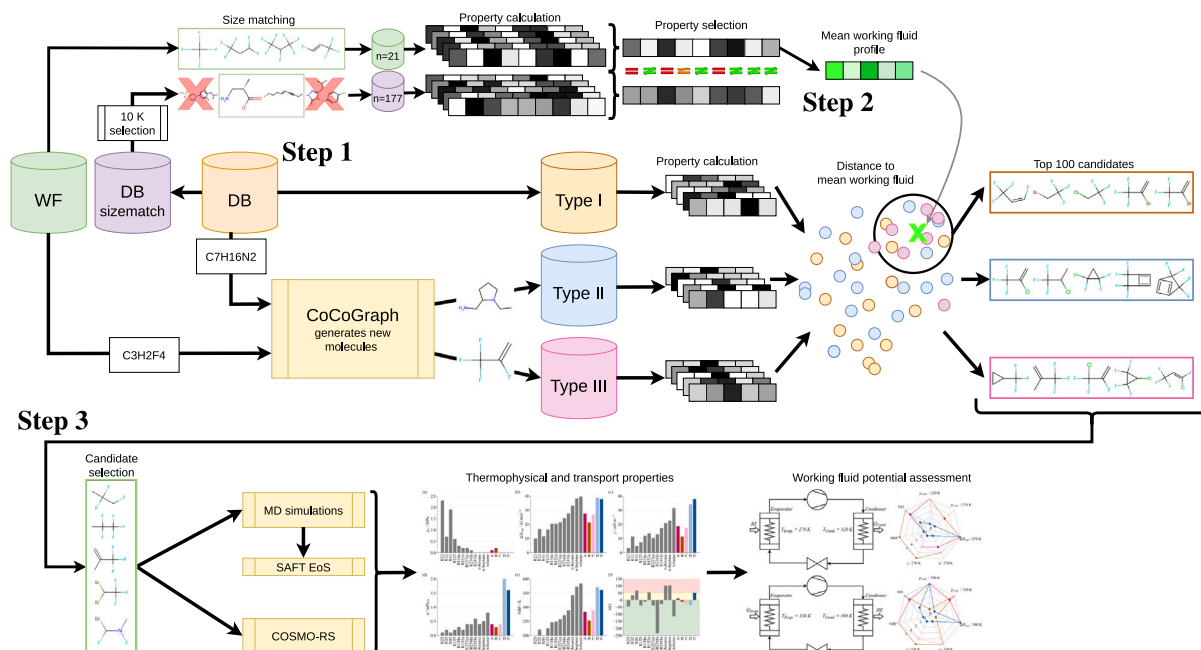


Fig. 1. Workflow illustration for the generative design and evaluation of novel working fluids. The workflow starts by establishing a physicochemical property profile for known working fluids and is followed by three main computational stages: **(Step 1)** Generation and property evaluation, wherein molecules are sampled from a reference database or generated via CoCoGraph, and then evaluated using RDKit to obtain a suite of physicochemical descriptors. **(Step 2)** Data distribution and filtering, which screens the candidate pools using distance-based filtering against the reference profile and visualizes them via t-SNE. **(Step 3)** Application feasibility and cycle performance, which uses MD, soft-SAFT, and process modelling to assess the top novel candidates for realistic refrigeration (VCRC) and heat pump (HTHP) applications.

and chemical environments covered by the available parameterization. The MD/soft-SAFT route adopted here is computationally more demanding but offers greater flexibility for generated molecules outside this applicability domain, and has been chosen to perform the thermophysical characterization of the selected fluids.

3. Results

3.1. Characterizing the physicochemical profile of working fluids

To identify new potential working fluid molecules, we first establish a comprehensive profile of the physicochemical properties that characterize working fluids. To do this, we compare a random sample of 10,000 molecules from the training set of CoCoGraph [20] to a curated set of 21 known working fluids (Fig. 2a). These 21 fluids are the size-matched subset (5–10 heavy atoms) of a broader list of 38 known working fluids reported in Table S1 of the Supplementary Information (SI). Because working fluids are relatively small molecules, typically containing between 5 and 10 heavy atoms, we filter our reference database to include only molecules within this range, resulting in 177 size-matched reference molecules for comparison (Fig. 2b). To ensure the statistical reliability of the selected descriptors and the resulting mean profile, we carry out extensive sensitivity analyses, including leave-one-out cross-validation and bootstrap resampling (see §S1.3 and Figs. S1–S2 in the SI).

For all molecules in our analysis — 21 working fluids and the 177 size-matched reference molecules —, we calculate 170 molecular descriptors using RDKit [48]. To identify which properties best distinguish working fluids from other molecules of similar size, we perform a statistical comparison. For each property, we calculate the Jensen–Shannon distance to quantify the difference between the probability distributions of the focal property for working fluids and size-matched reference molecules, which gives values ranging from 0 (identical distributions) to 1 (completely different distributions). Additionally,

we assess the statistical significance of these differences by performing non-parametric Mann–Whitney U tests, ensuring that the selected properties are significantly different between both groups ($p < 0.05$).

From this analysis, we identify 15 diverse properties that show the largest differences between working fluids and size-matched reference molecules (Fig. 2d). These properties span multiple molecular characteristics including electronic properties (maximum and minimum partial charge), surface area descriptors (Labute accessible surface area, molar refractivity, van der Waals surface area), topological features (number of heteroatoms, number of rotatable bonds), and structural complexity metrics (fingerprint density, molar refractivity). The analysis of these selected properties (Fig. 2c) reveals clear differences between working fluids and reference molecules, confirming their utility for identifying working fluid-like molecules. (For comparison, we also show properties where working fluids and reference molecules have similar distributions; Fig. 2e,f.) The mean values of the 15 selected properties across working fluids define a reference working fluid profile that serves as a target for candidate identification. This profile captures the essential physicochemical characteristics that distinguish working fluids from other small molecules, and enables us to search systematically for novel candidates with similar properties.

3.2. Identifying working fluid candidates through deep generative models and database search

With the established working fluid profile, we use three complementary strategies to identify potential new working fluid molecules (Fig. 3). First, we search CoCoGraph’s training database of 2.25 million molecules for those closely matching the working fluid profile (type I candidates). Second, we do the same with a database of 8.2 million novel molecules generated with CoCoGraph (type II candidates). Third, we use CoCoGraph to generate novel molecules with molecular formulas matching those of known working fluids, exploring structural variations while maintaining the same atomic composition (type III candidates). By design, all candidates generated by CoCoGraph (types

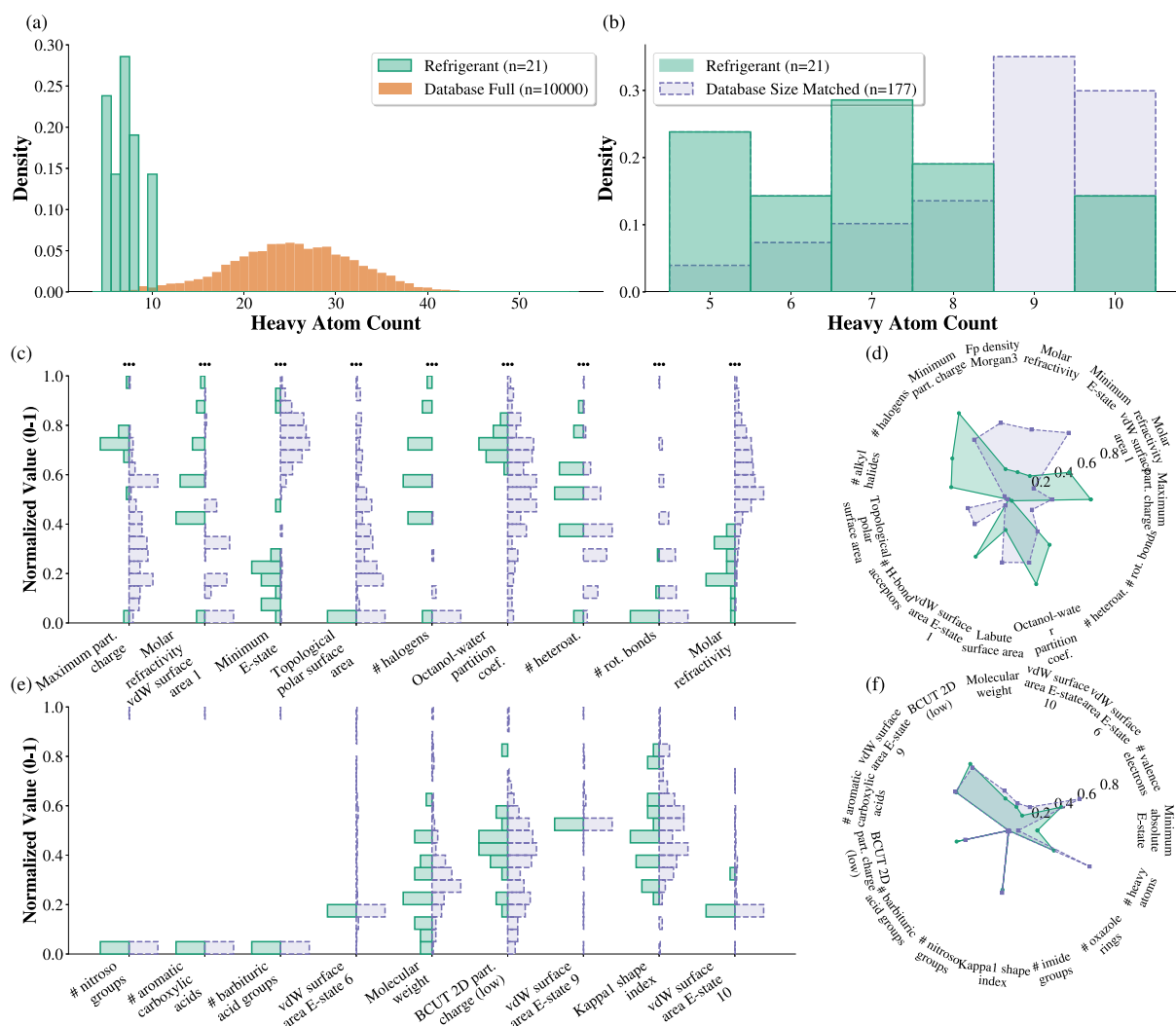


Fig. 2. Working fluid profiling against a size-matched molecular database. (a) Heavy-atom count distribution for 21 working fluids (green) versus 10,000 randomly sampled reference molecules (orange), highlighting differences in molecular size/composition. (b) Heavy-atom count distribution after restricting the database to the working fluid-relevant size window (5–10) used in subsequent analyses. Size-matched reference molecules are shown in purple. (c) Min–max normalized histograms for nine properties exhibiting significant distribution differences between working fluids and the size-matched database; significance is indicated by the panel annotations (***: $p < 0.001$). (d) Radar plot summarizing the 15 selected profiling descriptors, comparing the average property profile of working fluids versus the size-matched database. (e, f) Same as (c, d), but for descriptors showing no significant differences.

II and III) are strictly confined to its native 15-element vocabulary (H, B, C, N, O, F, Na, Mg, P, S, Cl, K, Ca, Br, and I). This constrained search space encompasses the elements most characteristic of organic working fluids while intentionally excluding rare, chemically unstable, or otherwise irrelevant elements (see §S1.4 of the SI for details). Further details regarding the generative model implementation, including the specific generation protocols for type II and III candidates, are also provided in §S1.4 of the SI.

To rank candidate molecules of all types according to their similarity to the working fluid profile, we calculate, for each candidate, the 15 key properties in the profile. We then compute the Euclidean distance between each candidate's property vector and the mean working fluid profile in the normalized 15-dimensional property space. This distance quantifies how similar each candidate is to the average working fluid across all relevant properties, with smaller distances indicating closer matches to the working fluid profile. Thus, from each of the three candidate groups (types I, II and III), we identify the 100 molecules with the smallest distances to this profile (Fig. 3a). The distance distributions reveal that candidates from all three sources include molecules that closely match the working fluid profile, with many candidates showing distances comparable to those of known working fluids.

To visualize the relationship between candidates and known working fluids in the 15-dimensional property space, we use t-SNE dimensionality reduction. In Fig. 3b, we show that the mean working fluid profile, known working fluids, and top candidates of types I, II and III cluster tightly together in a specific region of property space, in contrast to the broad cloud of all three types combined. This visually confirms that our search strategy successfully isolates molecules with physicochemical characteristics highly similar to those of known working fluids. It is important to note that while t-SNE effectively preserves local neighbourhood structure for visualization purposes, the actual candidate ranking is strictly determined by the 15-dimensional Euclidean distance to the mean profile, rather than by the geometric distances apparent in this 2D projection. To confirm the candidate selection, we further analyse specific properties. For example, distributions of Labute accessible surface area and molar refractivity (Fig. 3c,d,e) show that top candidates from all three types closely match the working fluid profile, with distributions that overlap substantially with those of known working fluids and differ from other molecules generated in the same way. Finally, we examine the top 10 candidates of each type using radar plots comparing their property profiles to both the mean working fluid profile and the mean profile of size-matched reference

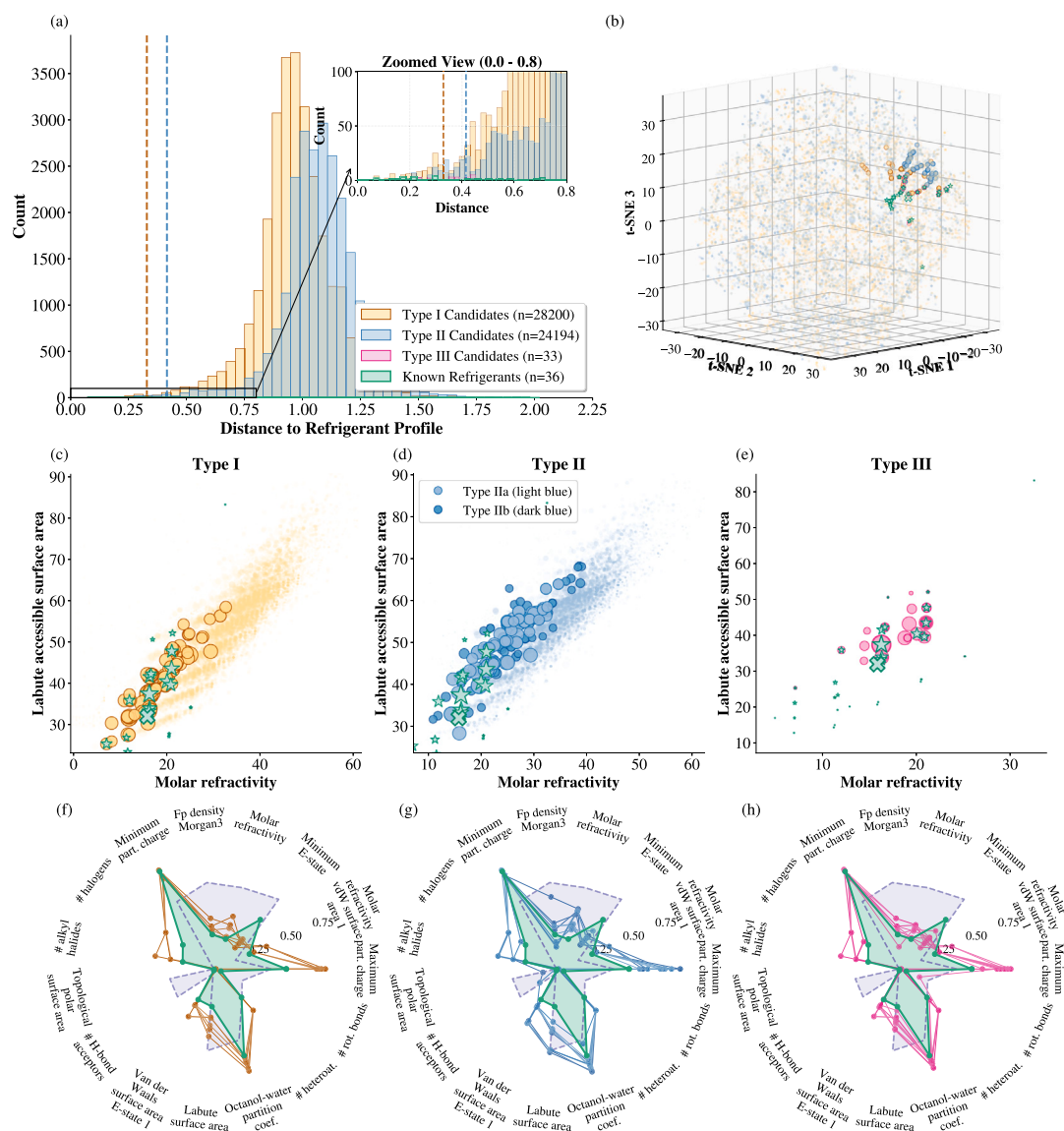


Fig. 3. Profile-based candidate screening for working fluid-like molecules across putative candidates. (a) Distribution of distances to the working fluid property profile for type I molecules (yellow), type II molecules (blue), and type III molecules (pink) (see text for definitions of each type). The zoomed inset shows the region of shortest distances and the selection of top candidates. (b) t-SNE projection of the profiling descriptor space, showing known working fluids (green stars, with the green cross representing the mean profile), and selected candidates of each type. (c)–(e) Scatter plots of two representative physicochemical descriptors (molecular refractivity and Labute accessible surface area) for the working fluid reference, known working fluids, and candidates of each type; marker size encodes overall distance to the mean working fluid profile. In panel (d) light blue markers indicate top candidates from the CoCoGraph-generated database (type IIa), while dark blue markers highlight an additional set of novel candidates verified to be absent from PubChem (type IIb). (f)–(h) Radar plots of the 15 profiling descriptors comparing the mean working fluid profile, the size-matched database baseline (purple dashed line), and top candidates from each type, illustrating the closeness of candidate profiles to the mean working fluid target.

molecules (Fig. 3f,g,h). These comparisons demonstrate that top candidates consistently match the working fluid profile more closely than the database profile across multiple properties, confirming their potential as alternative working fluids.

After this verification step, we proceed to select candidates for further characterization. For type II candidates, we distinguish between: (i) molecules that, despite being absent from CoCoGraph's training set (the model has never seen them at any stage, and are therefore novel molecules from its perspective), exist in PubChem (type IIa); and (ii) molecules that are absent from PubChem (type IIb). We then select 100 molecules of types I, IIa, IIb and III, and filter out the 21 size-matched known working fluids used to construct the profile (Table S1) and/or duplicates, which leads to candidate sets of 65 type I molecules, 56 type IIa molecules, 84 type IIb molecules, and 23 type III molecules. The top-ranked candidates from each of these four filtered

pools are detailed in Tables S2–S5 of the SI. To confirm the robustness of these candidate rankings, we conduct further sensitivity analyses, including leave-one-out cross-validation of the mean profile, descriptor ablation, and evaluation under alternative distance metrics (see §S1.3, Figs. S3–S5, and Tables S6–S11 in the SI). From these pools, we select five compounds (A to E) for detailed thermodynamic characterization (Table 1). To select these compounds, we apply a series of exclusion criteria based on application-oriented considerations, while also favouring simpler molecular structures. Molecular simplicity was considered both with regard to the potential synthesis of the proposed compounds and to modelling feasibility. In particular, since our MD simulations rely on a general and transferable force field, the parameterization of increasingly complex molecular structures may introduce additional challenges. The applied exclusion criteria are the following: (i) We exclude ionic species to restrict the analysis to neutral compounds; (ii)

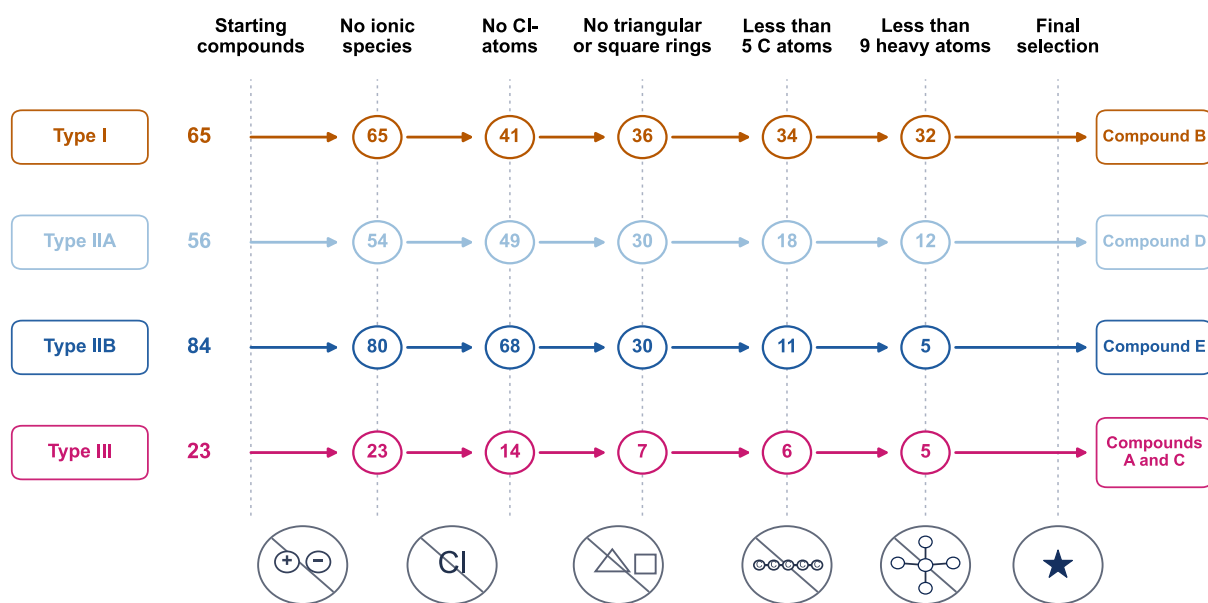


Fig. 4. Compound selection flowsheet. Screening flowchart detailing the number of candidates of each type that remained available after each exclusion criteria.

we omit chlorine-containing molecules due to environmental restrictions associated with ODP; and (iii) we exclude 3-carbon and 4-carbon rings due to their relatively low stability; and we give preference to compounds with limited number of atoms to avoid excessively high boiling points, by only considering compounds with (iv) less than five carbons and (v) less than nine heavy atoms. Fig. 4 summarizes the selection process, showing the number of compounds of each type that remain after applying each filter.

Table 1 lists the five compounds selected for subsequent thermodynamic characterization: (A) 1,2,2-trifluoropropane; (B) 1,1,1,2,2-pentafluoropropane; (C) 3,3,3-trifluoro-2-methylpropene; (D) 1,1-dibromo-2,2,2-trifluoroethane; and (E) 1-bromo-1-fluoro-N-fluoro-N-methylmethanamine. To accurately contextualize the discovery aspect of our workflow, we classify these candidates across distinct novelty tiers (Table 1). Compound B is a type I candidate (already present in the training database and PubChem) and is a known substance in the refrigerant literature (designated as R-245cb). Rather than discarding it, we explicitly retain compound B as a critical validation benchmark, leveraging its available experimental vapour–liquid equilibrium densities and vapour pressures [49–52] to ground our subsequent thermodynamic simulations. In contrast, compounds A, C (both type III), and D (type IIA) are novel to the generative model (absent from its training database) and are absent from the refrigerant literature, despite having existing PubChem records; consequently, little to no thermophysical data are available for these species. Finally, compound E (type IIB) is a genuinely new chemical entity, being completely absent from both the model’s training data and the PubChem database.

3.3. Thermodynamic modelling of candidate working fluids

We evaluate the five candidates using multiscale thermodynamic models. The details of the methodologies employed are provided in §§1.5–S1.7 of the SI. First, we obtain the σ -profiles of the selected compounds from COSMO-RS and compare them with those of established working fluids, specifically HFCs and alkanes (Fig. S8 in the SI). The σ -profiles indicate that the molecular surface polarity can be broadly divided into three distinct regions. Negative charge densities ($\sigma < -0.005 \text{ e.}\text{\AA}^{-2}$) correspond to positively polarized surface segments with hydrogen-bond donating character, whereas positive charge densities ($\sigma > +0.005 \text{ e.}\text{\AA}^{-2}$) are associated with negatively polarized regions exhibiting hydrogen-bond accepting behaviour. Charge densities in the

intermediate range ($-0.005 \leq \sigma \leq +0.005 \text{ e.}\text{\AA}^{-2}$) are associated with essentially neutral, non-polar surface areas of the molecules. The σ -profiles of compounds A, B, and C reveal a systematic effect of fluorine substitution. An increasing number of fluorine atoms leads to a shift of the left-side peak towards more negative values, reflecting enhanced positive polarization of the hydrogen atoms, while the negative polarization of the fluorine atoms becomes less pronounced, subsequently driving the positive peaks to the left as well. This behaviour is consistent with the trend observed in the σ -profiles of HFCs. Concerning compounds D and E, their σ -profiles exhibit peaks at more negative charge densities ($-0.015 \leq \sigma \leq -0.010 \text{ e.}\text{\AA}^{-2}$), comparable to those observed for highly fluorinated species such as R-23, R-125, and R-227ea. Here, these features can be attributed to the presence of bromine atoms. The high electronegativity of bromine enhances the polarization of neighbouring hydrogen atoms, thereby increasing their hydrogen-bond donating character.

Therefore, when compared qualitatively with established working fluids, the similarities in specific σ -profile regions suggest that the proposed compounds may share aspects of intermolecular interaction behaviour with established working fluids and, consequently, comparable thermophysical properties. Based on this observation, we use a multiscale computational approach to assess key properties of the selected compounds, including vapour–liquid equilibrium densities, vapour pressures, enthalpies of vaporization, surface tensions, saturated-liquid self-diffusion coefficients, and viscosities. First, we obtain these quantities from MD simulations and then use the results as reference data for the development and parameterization of a thermodynamic soft-SAFT model (Table 2). As an additional test, we include COSMO-RS predictions of saturated-liquid transport properties for comparison.

To assess the predictive reliability of the present methodology, we first compare our predictions with the experimental vapour–liquid equilibrium data for compound B (R-245cb) identified in the literature. As an additional validation, we apply the same computational workflow to an established working fluid, the HFO R-1234yf, for which experimental data are available in the literature for a range of relevant thermophysical properties, including vapour–liquid equilibrium densities, vapour pressures, surface tensions, and viscosities of the saturated liquid phase. Overall, the MD predictions and the soft-SAFT adjustments show very good agreement with the available experimental data for both compounds, while the COSMO-RS predictions of viscosity

Table 1

Candidate compounds identified and modelled in this work. Code, formula, name, SMILES, candidate type, novelty status, and molecular weight. Candidate types refer to: database search (I), novel generation present in PubChem (IIa), novel generation absent from PubChem (IIb), and molecular formula-matched generation (III).

Compound	Code	Formula	Name	SMILES	Type	Novelty status	$M_w/\text{g.mol}^{-1}$
	A	$\text{CH}_3\text{CF}_2\text{CH}_2\text{F}$	1,2,2-trifluoropropane	<chem>CC(F)(F)CF</chem>	III	Absent from refrigerant literature	98.07
	B	$\text{CH}_3\text{CF}_2\text{CF}_3$	1,1,1,2,2-pentafluoropropane	<chem>CC(F)(F)C(F)(F)F</chem>	I	Known in literature (R-245cb)	134.05
	C	$\text{CH}_3\text{CCH}_2\text{CF}_3$	3,3,3-trifluoro-2-methylpropene	<chem>C=C(C)C(F)(F)F</chem>	III	Absent from refrigerant literature	110.08
	D	$\text{CF}_3\text{CBr}_2\text{H}$	1,1-dibromo-2,2,2-trifluoroethane	<chem>FC(F)(F)C(Br)Br</chem>	IIa	Absent from refrigerant literature	241.83
	E	CH_3NFCFBr	1-bromo-1-fluoro-N-fluoro-N-methylmethanamine	<chem>CN(F)C(F)Br</chem>	IIb	Genuinely new chemical entity	159.96

Table 2

List of molecular parameters used in this work. This includes polar soft-SAFT, DGT, and FVT parameters, together with deviations for the fitted properties. The parameters are described in §S1.7 of the SI.

Compound	m	$\sigma/\text{\AA}$	$\epsilon k_B^{-1}/\text{K}$	$\mu \cdot 10^{30}/\text{C.m}$	x_p	AAD $_p/\%$	AAD $_{\rho}/\%$	$c_{ii} \cdot 10^{19}$	AAD $_{\text{DGT}}/\%$	α^a	$B \cdot 10^2$	$L_v/\text{\AA}$	AAD $_{\text{FVT}}/\%$
A	2.441	3.616	189.6	6.12	0.60	10.31	1.50	1.7	13.5	18.958	3.308	1.866	4.62
B	2.018	3.975	151.1	7.95	0.63	1.48	1.06	2.0	4.5	16.283	3.413	1.514	1.90
C	2.021	4.049	203.3	8.23	0.55	4.57	0.34	2.2	14.1	17.698	3.054	2.108	0.45
D	2.242	3.954	300.6	4.69	1.00	3.07	0.47	3.2	7.4	14.561	2.554	1.522	0.51
E	2.548	3.632	287.3	3.32	0.85	3.82	0.49	3.0	10.5	16.144	3.268	1.827	1.57

^a α units: $\text{J.m}^3\text{.mol}^{-1}\text{.kg}^{-1}$.

should be interpreted primarily in qualitative terms (Figs. S6 and S7 in the SI; see Table S12 in SI for the corresponding soft-SAFT model parameters). Nevertheless, the inherent limitations associated with the use of a general and transferable OPLS-AA force field in our MD simulations should be acknowledged. Since the force field was not specifically parameterized for the compounds investigated, deviations from experimental values are expected, particularly when quantitative accuracy is compared with that achievable using compound-specific force fields [35]. Despite these limitations, our MD simulations reproduce the overall thermophysical behaviour of these two fluids and provide a qualitatively consistent description of their properties. Taken together, these results provide an independent assessment of the predictive reliability of the workflow and validate its applicability as a preliminary screening tool for novel working fluids, with the aim of estimating and comparing thermodynamically relevant properties, identifying promising candidates, and guiding subsequent, more detailed investigation and experimental verification.

In Fig. 5, we display the results obtained from our computational workflow for the five candidate working fluids A–E. Table 2 lists the parameters of the thermodynamic soft-SAFT model. The polar soft-SAFT framework allows accurate reproduction of the vapour–liquid equilibrium densities and vapour pressures obtained from MD simulations for all compounds, thereby extending the applicability of the model up to the critical point. Although the enthalpy of vaporization was not included in the fitting procedure, both soft-SAFT and MD yield comparable values, indicating internal consistency between models (Fig. 5(c)). After fitting the DGT influence parameter (c_{ii}), the modelled surface tensions show good agreement with simulation data (Fig. 5(d)), with a maximum absolute average deviation (AAD) of 14.1% for compound C (Table 2). Likewise, the integration of FVT within the soft-SAFT framework provides an excellent description of the viscosities obtained from MD simulations, with a maximum AAD of 4.62% for compound A (Fig. 5(f)). The AAD values reported in Table 2 quantify the agreement

between the model predictions and the MD data used for parameterization, but should not be interpreted as confidence intervals for the fitted parameters. As in other SAFT-type models, correlations among parameters such as m , σ , and ϵ may result in different parameter combinations providing comparably satisfactory descriptions of the fitted properties. The reported parameter sets should therefore be regarded as effective values associated with the adopted fitting protocol, rather than as uniquely determined molecular descriptors. Additionally, the description of transport properties using COSMO-RS remains mostly qualitative, as we used this quantum-chemistry-based approach without fitting any specific parameters.

In Fig. 6, we compare the estimated thermophysical properties of the selected compounds with the properties of established working fluids, including HFCs and hydrocarbons used in refrigeration and heat pump applications, at 270 K, along with their normal boiling points (see Table S13 in SI for data available in the literature for the 21 reference fluids used to construct the working fluid profile). The estimated properties for compounds A, B, and C are broadly comparable to those of hydrofluoropropanes such as R-227ea, R-236fa, and R-245fa. In contrast, compounds D and E have estimated vapour pressures, normal boiling points, and enthalpies of vaporization that fall within the range of hydrocarbons such as *n*-hexane, *n*-heptane, and toluene, albeit with generally higher surface tensions and significantly higher viscosities. Fig. S9 in the SI shows the temperature dependence of the thermophysical properties of the five selected compounds and of established working fluids. This comparison further highlights the overlap in thermophysical behaviour between the selected compounds and the benchmarked systems across a broad temperature range. Additionally, we predict the NFI of each candidate, using the σ -profiles derived from COSMO calculations as inputs in the neural network model developed by Albà et al. [14]. Further details on the NFI predictions are provided in §S1.8 of the SI. The resulting values indicate that compounds B, C, and D are classified as non-flammable (A1), while

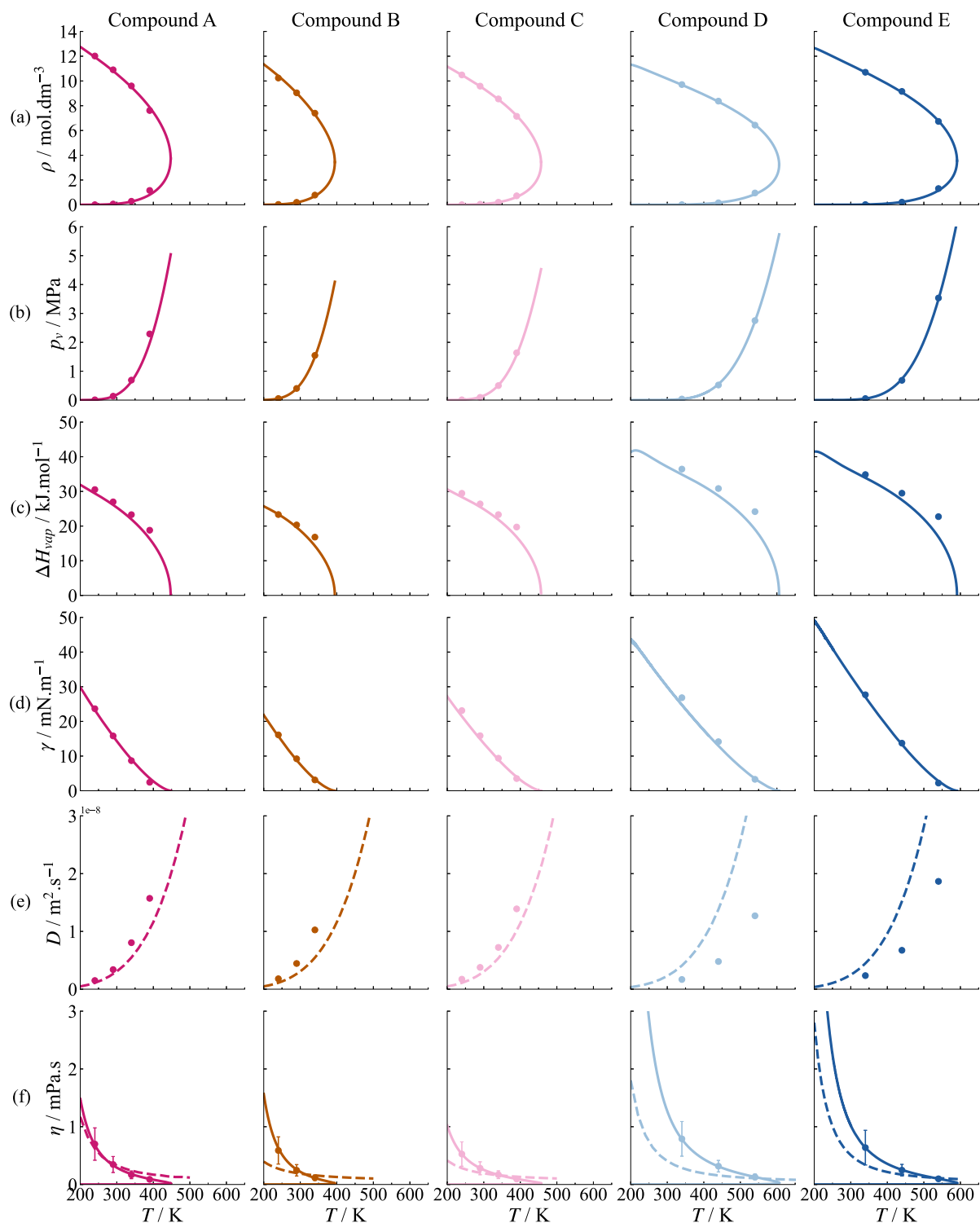


Fig. 5. Thermodynamic modelling of candidate working fluids. For each of the five candidate compounds, we show: (a) vapour–liquid coexistence density; (b) vapour pressure; (c) enthalpy of vaporization; (d) surface tension; (e) saturated liquid self-diffusion coefficient; and (f) saturated liquid viscosity. Circles are MD simulation results; solid lines are polar soft-SAFT fits, and dashed lines are COSMO-RS results. Error bars are shown for MD simulation results where the uncertainty exceeds the symbol size.

compound A, which exhibits an NFI of 12.2, falls within the mildly flammable category (A2L), and compound E, which yields an NFI of 52, slightly exceeds the flammability threshold (A2) (Fig. 6(f)). Note that the safety assessment performed in this work is limited to predicted flammability and should not be interpreted as a detailed evaluation of the safety of the proposed compounds. Other relevant aspects, including toxicity, chemical and thermal stability, reactivity, and decomposition

products, are beyond the scope of the present computational screening and would require dedicated experimental and/or computational assessment before these compounds could be considered for practical implementation. Additionally, we calculated the GWP of the five compounds using the artificial neural network developed by Devotta et al. [54]. The predicted values were GWP(A)=945, GWP(B)=4000, GWP(C)=0.1, GWP(D)=1395, and GWP(E)=0.1. Compounds A, B, and

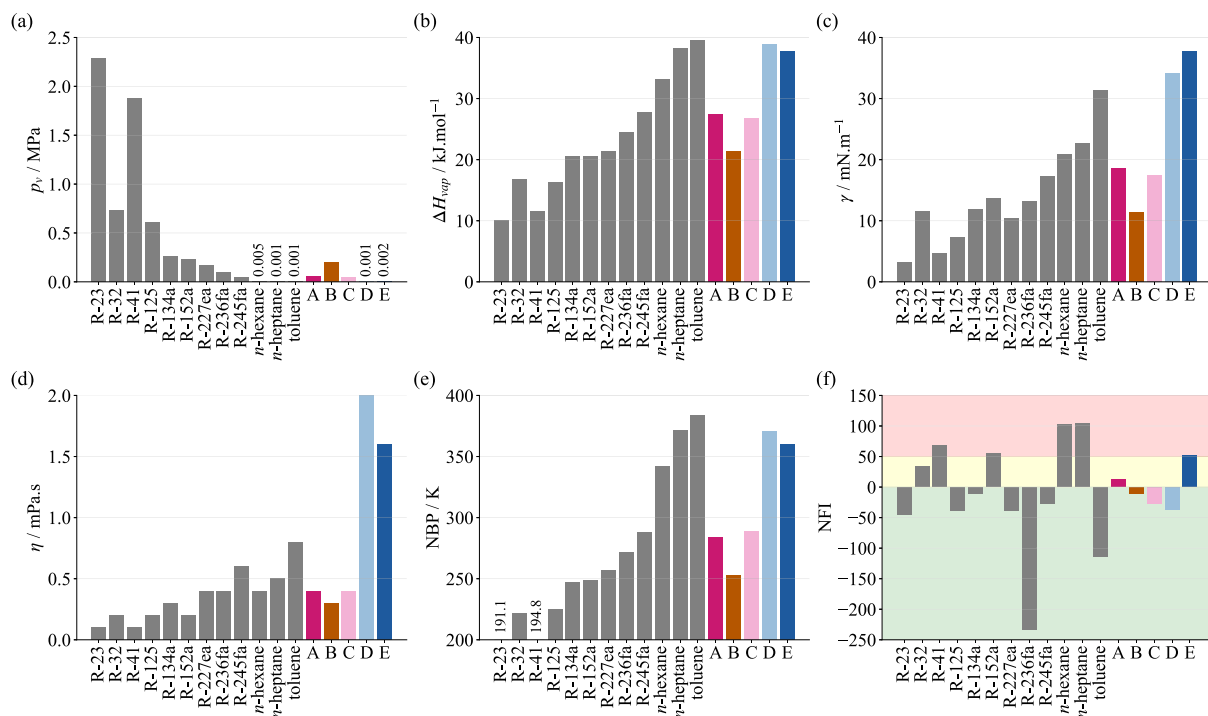


Fig. 6. Comparison of thermodynamic properties of candidate working fluids to established working fluids. (a) Vapour pressure. (b) Enthalpy of vaporization. (c) Surface tension. (d) Viscosity of the saturated liquid at 270 K. (e) Normal boiling point (NBP). (f) Normalised flammability index (NFI). Data in panels (a) to (e) for established working fluids were taken from the NIST Chemistry WebBook database [53], whereas for the five selected compounds were obtained from soft-SAFT predictions. NFIs were predicted using the neural network model of Albà et al. [14].

D exhibit relatively high predicted GWP values, all considerably exceeding the EU regulatory limit of 150 [11]. In contrast, the very low values obtained for compounds C and E suggest that these compounds may be promising working fluids, also from an environmental perspective. In particular, compound C stands out for combining a very low predicted GWP and flammability, with favourable thermodynamic properties, further supporting its potential as a novel working fluid. It should be noted, however, that the predicted GWP values are preliminary estimates and should, therefore, be interpreted with caution. A comprehensive evaluation of the environmental impact of these compounds would require the assessment of additional criteria, including ODP and atmospheric lifetime, which are beyond the objectives of the present work.

3.4. Benchmarking analysis of the candidate working fluids

Finally, we evaluate the suitability of the candidate compounds to replace conventional working fluids in VCRCs and HTHPs. To do so, we compare a set of their key physicochemical properties with those of R-134a, a benchmark fluid for VCRCs, and R-245fa, a benchmark for HTHPs [55]: vapour pressure, NBP, enthalpy of vaporization, viscosity, and surface tension, obtained from our polar soft-SAFT framework fits, together with the predicted NFIs. For this, we consider representative VCRC and HTHP configurations, both consisting of a single-stage compressor, condenser, evaporator, and thermostatic expansion valve (TXV) (see Fig. S10 in SI). In both configurations, the cycle begins with the isentropic compression of the working fluid, followed by heat rejection in the condenser, where the superheated vapour releases latent heat to the surroundings and reaches saturation at the condenser pressure. In the case of the HTHP, this stage corresponds to the useful heat delivery process. The saturated liquid is subsequently expanded through the TXV, producing a two-phase vapour–liquid mixture, while regulating the refrigerant mass flow rate. The mixture then enters the evaporator, where isobaric vaporization occurs, and the working fluid reaches the saturated vapour state. In the VCRC, this phase transition

provides the desired refrigeration effect by absorbing heat from the environment. For the VCRC, the evaporator and condenser temperatures are fixed at 270 K and 320 K, respectively, whereas for the HTHP they are set to 330 K and 390 K, corresponding to typical operating conditions [56,57].

The performance of refrigeration and heating cycles strongly depends on the thermophysical properties of the working fluid, which must satisfy a set of well-established criteria. Specifically: (i) a low NBP is desirable to facilitate vaporization; (ii) the vapour pressure in the condenser should be as low as possible, while (iii) the vapour pressure in the evaporator should be as high as possible to reduce the required compression power and the consequent operating costs; (iv) vapour pressure must remain above atmospheric pressure to avoid working under vacuum operations, due to the associated technical and energetic penalties; (v) both viscosity and surface tension should be as low as possible, as these properties directly affect heat transfer and pumping efficiency; (vi) the enthalpy of vaporization should be maximized, as it determines the amount of heat that can be transferred; and (vii) a low NFI is preferable. With these criteria in mind, we compare vapour pressures at both the condenser and evaporator temperatures, while viscosity and surface tension are evaluated at the lowest operating temperature of each cycle (the evaporator temperature), representing the worst case scenario. Finally, we compare the enthalpy of vaporization at the temperature of the primary heat exchange component in each system, namely the evaporator for the VCRC and the condenser for the HTHP. Then, we assign performance scores on a five-level scale (1 being the least favourable, 5 the most favourable), with the detailed scoring criteria provided in Table S14 in the SI.

Fig. 7 summarizes the resulting performance scores for each compound across both VCRC and HTHP configurations. For the VCRC, compound B has the highest performance scores, with predicted properties closely comparable to those of R-134a. Unfortunately, its high GWP prevents its use as a refrigeration fluid. The remaining compounds are excluded from VCRC applications primarily due to vapour pressures below atmospheric pressure at the evaporator and/or excessively high

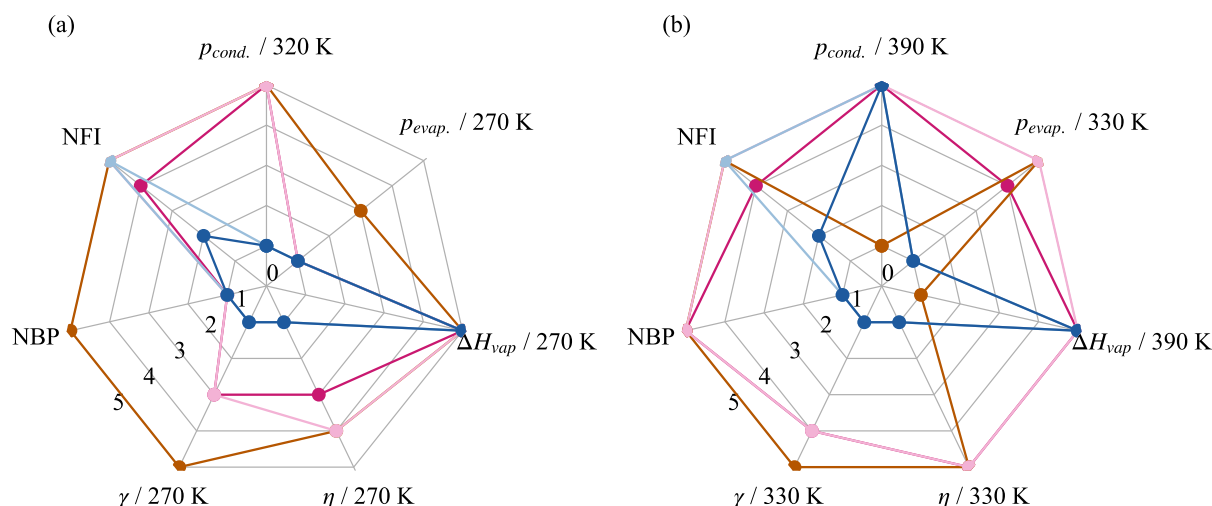


Fig. 7. Benchmarking analysis of the candidate working fluids. We benchmark the five candidate compounds A (●), B (●), C (●), D (●), and E (●) against two reference working fluids: (a) R-134a and (b) R-245fa. The benchmark is in terms of seven key physicochemical and safety properties: vapour pressures at the condenser and the evaporator, enthalpy of vaporization, viscosity, surface tension, NBP, and NFI. Performance scores are reported on a five-level scale (1=least favourable, 5=most favourable) based on the similarity to the properties of the reference fluids in VCRCs and HTHPs, respectively for (a) and (b). Scoring criteria are detailed in Table S14 of the SI.

NBPs. In the context of the HTHP, compounds A and C achieve the highest rankings. However, compound C clearly stands out, attaining the maximum score for six of the seven evaluated properties, with estimated values comparable to, and in some cases exceeding, those of R-245fa. By contrast, compound B is unsuitable for HTHP operations due to its excessively high condenser vapour pressure, while compounds D and E display consistently poor performance, mainly as a result of their high NBPs and viscosities. These results confirm the technical suitability of compound B as a working fluid of VCRCs and indicate that compounds A and C have properties compatible with their use in HTHPs. Among them, compound C emerges as the most promising candidate.

This screening framework provides a comparative basis for identifying promising working-fluid candidates by combining thermophysical performance with additional criteria such as flammability and GWP. We use the thermophysical analysis to establish whether the predicted properties of the selected compounds fall within ranges associated with established working fluids and to identify similarities in their overall behaviour. However, it must be noted that, given the uncertainties inherent to the individual predictive methods, we interpret these comparisons as indicative of relative suitability rather than as exact quantitative distinctions between candidates.

To complete the property-based screening, we subsequently examine the leading candidates at the process level to assess whether their predicted thermophysical characteristics are translated into competitive cycle performance under representative operating conditions. Compound B, corresponding to R-245cb, is assessed in VCRC, whereas compound C is evaluated in a HTHP cycle, as described in the beginning of this Section. For these calculations, we assume steady-state operation, negligible pressure drops and heat losses, isenthalpic expansion, and a compressor isentropic efficiency of 85%. Additionally, we allow a superheating of 15 K and 30 K in the evaporator for VCRC and HTHP, respectively, as well as a subcooling of 5 K in the condenser for both cycles. The relatively large superheating employed in the HTHP ensures dry compression. All calculations are done for a cooling (VCRC) or heating (HTHP) capacity of 10 kW, being performed with the soft-SAFT EoS.

We have evaluated a series of process performance indicators and have compared them with the results obtained using benchmark refrigerants. The refrigeration and heating coefficients of performance (COP), volumetric cooling and heating capacity (VCC and VHC), specific refrigeration and heating effect, pressure ratio, condenser pressure,

discharge-line temperature, and suction density were calculated as described in §S5 of the SI. Table 3 lists the performance indicators obtained for the selected working fluids and benchmark refrigerants in the VCRC and the HTHP cycles.

The COP predicted for compound B (2.10) is close to that calculated for the R-134a (2.24) under the same cycle operating conditions, corresponding to a difference of approximately 6%. This direct comparison therefore indicates that compound B can achieve refrigeration-cycle performance comparable to that of R-134a. The pressure ratio is also very similar (9.71 versus 9.32), while compound B exhibits a lower discharge-line temperature. In contrast, its volumetric cooling capacity (VCC) is lower than that of R-134a (0.62 versus 0.91 kJ.L⁻¹), indicating that comparable COP does not necessarily imply equivalent volumetric performance. The cycle results nevertheless confirm that compound B has technically suitable thermodynamic behaviour for VCRC. However, as noted before, its high reported GWP prevents its practical use. Consequently, this result should be regarded primarily as a validation of the proposed multiscale screening methodology rather than as a recommendation for its practical adoption as a sustainable replacement for R-134a.

For the HTHP application, compound C operates between 0.153 and 1.720 MPa, giving a pressure ratio of 11.21. Despite the relatively large temperature lift, the calculated discharge temperature is 405.4 K, approximately 10 K above the condensation temperature. Compound C provides a heating effect of 19.47 kJ.mol⁻¹ and a volumetric heating capacity of 1.10 kJ.L⁻¹. The resulting heating COP is 2.56. The direct comparison with R-245fa under the same operating conditions shows very similar cycle performance (Table 3). Compound C exhibits a slightly higher COP, while the pressure ratio, discharge-line temperature, and volumetric heating capacity remain comparable to those of R-245fa. These results indicate that compound C can achieve a good HTHP cycle performance under the investigated operating conditions and reinforces its selection as the most promising candidate for heating applications, although experimental evaluation of its safety, stability, and environmental properties remains necessary before practical implementation.

4. Conclusions

In this work, we introduced a hybrid multiscale computational workflow that integrates deep generative learning with molecular simulation, COSMO-RS calculations, soft-SAFT modelling, preliminary environmental and safety screening, and cycle-level evaluation to identify

Table 3

Process-level performance indicators for the selected working fluids and benchmark refrigerants.

Parameter	VCRC		HTHP	
	Compound B	R-134a	Compound C	R-245fa
Condenser pressure, MPa	1.06	1.33	1.72	1.89
Pressure ratio	9.71	9.32	11.21	11.84
Discharge-line temperature, K	334.5	352.8	405.4	406.7
Suction density, kg.m ⁻³	6.73	6.75	6.36	8.06
Specific refrigeration/heating effect, kJ.mol ⁻¹	12.42	13.75	19.47	19.42
Volumetric cooling/heating capacity, kJ.L ⁻¹	0.62	0.91	1.10	1.17
Coefficient of performance (COP)	2.1	2.24	2.56	2.47

novel working fluids. We combined the ability of generative models to explore previously unconsidered regions of molecular space with physically grounded methods for assessing the thermophysical and process-relevant performance of the resulting candidates.

We explored three complementary regions of chemical space. First, we retrieved type I candidates from the 2.25 million molecules contained in the CoCoGraph training database. Second, we selected type II candidates from 8.2 million novel molecules generated by CoCoGraph. Third, we generated type III candidates by constraining CoCoGraph to the molecular formulas of known working fluids, thereby exploring new molecular structures, while retaining working-fluid-relevant atomic compositions. Together, these strategies enabled us to identify chemically diverse candidates, ranging from saturated fluorinated hydrocarbons to unsaturated, brominated, and nitrogen-containing halogenated structures.

We then screened regions of chemical space against the 15-dimensional physicochemical profile of established working fluids. Using successive application-oriented filters, we excluded ionic species, chlorine-containing compounds, small strained rings, and molecules likely to exhibit excessively high boiling points. This process yielded five candidates representing the different regions explored. We predicted their thermophysical properties using MD simulations and COSMO-RS calculations and modelled the results with the soft-SAFT EoS. Additionally, we estimated their GWP and flammability using property-specific neural-network models.

Finally, we performed cycle simulations to evaluate the most promising compounds at the process level. 1,1,1,2,2-Pentafluoropropane (Compound B) achieved a refrigeration COP of 2.10, approximately 6% below that of R-134a under the same operating conditions. Although its high predicted GWP prevents us from recommending it as a sustainable replacement, this result validates the ability of our workflow to identify molecules with suitable cycle performance. More broadly, these results should be considered in the context of the limited chemical space available for refrigeration applications. The low boiling points required for refrigeration, along with the other physicochemical, structural, environmental, and application requirements, impose strict constraints on the molecular structures that can serve as suitable working fluids. Nevertheless, only a small subset of the molecules generated by CoCoGraph was subjected to the complete multiscale computational workflow, owing to the computational cost associated with the successive stages of the methodology. The investigation of the structures excluded in future work could provide a broader assessment of the chemical space identified by CoCoGraph and further explore the potential of the generated candidates.

For heat-pump applications, by comparison, these constraints are less restrictive, resulting in a substantially larger pool of potential candidates. This aspect is exemplified by one of the main outcomes of this work, the identification of 3,3,3-trifluoro-2-methylpropene (Compound C) as a promising candidate for heat-pump cycles. Compound C exhibited good performance for HTHP applications, with a heating COP of 2.56, slightly exceeding that calculated for R-245fa. The predicted pressure ratio, discharge temperature, and volumetric heating capacity were also comparable to those of the benchmark fluid. In combination with a very low GWP, favourable thermodynamic and transport

properties, and low flammability, Compound C stands out as the best candidate in this study.

We emphasize that our results provide a prioritization for targeted experimental investigation rather than a definitive ranking of deployable working fluids. In the present assessment, we limited the environmental and safety evaluation to predicted GWP and flammability. Future studies should examine toxicity, chemical and thermal stability, reactivity, exposure risks, and potential decomposition products. They should also experimentally validate the predicted properties and systematically propagate model uncertainties through the complete workflow. Moreover, although our open-loop framework is designed to screen candidates by similarity to established working fluids rather than to optimize for superior performance, extending it towards closed-loop molecular optimization [58,59] will require computationally efficient and well-calibrated multifidelity surrogate models capable of approximating the high-fidelity molecular-simulation and thermodynamic-modelling stages [60].

In summary, we establish a novel and transferable route from generative molecular design to process-relevant performance evaluation. By systematically exploring complementary regions of chemical space, substantially narrowing the number of plausible structures, and identifying candidates worthy of experimental validation, our framework represents an important step towards the computationally guided development of next-generation working fluids.

CRedit authorship contribution statement

Manuel Ruiz-Botella: Formal analysis, Methodology, Visualization, Software, Investigation, Writing – original draft. **Tiago M. Eusébio:** Formal analysis, Methodology, Visualization, Data curation, Software, Investigation, Writing – original draft. **Marta Sales-Pardo:** Formal analysis, Methodology, Visualization, Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Roger Guimerà:** Formal analysis, Methodology, Visualization, Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Sabrina Belén Rodríguez-Reartes:** Formal analysis, Methodology, Visualization, Conceptualization, Supervision, Writing – review & editing. **Fèlix Llorell:** Formal analysis, Methodology, Visualization, Conceptualization, Supervision, Writing – review & editing, Funding acquisition.

Acknowledgements

This research was supported by projects PID2023-149713OB-I00 and PID2022-142600NB-I00 (F.L., M.S.-P. and R.G.) from MCIN/ AEI/ 10.13039/ 501100011033, and by the EU Next Generation, URV and SEPE (2022PMF-INV-10) (M.R.-B.). We thankfully acknowledge the computer resources at MareNostrum5 and the technical support provided by BSC (RES-IM-2025-1-0021).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

SI contains results referenced in text. Methods are described in SI.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ccej.2026.183129>.

Data availability

The deep generative model CoCoGraph used to produce the initial molecular spaces is open source and available at <https://github.com/manurubo/cocograph>. To ensure full reproducibility of the downstream screening workflow, a complete data and code repository containing the candidate generation scripts, descriptor computation, ranking procedures, filtering steps, and the final top-candidate lists presented in this work is openly available at <https://github.com/manurubo/refrigerant-screening-workflow>.

References

- [1] M. Baha, S. Hammami, J.L. Supont, *The Role of Refrigeration in the Global Economy* (3rd edition), Technical Brief, International Institute of Refrigeration (IIR), Paris, France, 2025.
- [2] P. Giménez-Prades, J. Navarro-Esbrí, C. Arpagaus, A. Fernández-Moreno, A. Mota-Babiloni, Novel molecules as working fluids for refrigeration, heat pump and organic Rankine cycle systems, *Renew. Sustain. Energy Rev.* 167 (2022) 112549, <https://doi.org/10.1016/j.rser.2022.112549>.
- [3] A.K. Vuppaladadiyam, E. Antunes, S.S.V. Vuppaladadiyam, Z.T. Baig, A. Subiantoro, G. Lei, S.-Y. Leu, A.K. Sarmah, H. Duan, Progress in the development and use of refrigerants and unintended environmental consequences, *Sci. Total Environ.* 823 (2022) 153670, <https://doi.org/10.1016/j.scitotenv.2022.153670>.
- [4] R. Yıldırım, A. Akyüz, K. Kumaş, A.D. Tuncer, A. Gungor, A comprehensive overview of refrigerants from the past to the modern day: energy, exergy, and environmental perspectives, *J. Therm. Anal. Calorim.* 150 (5) (2025) 3201–3223, <https://doi.org/10.1007/s10973-025-14050-w>.
- [5] M.O. McLinden, M.L. Huber, (R)evolution of refrigerants, *J. Chem. Eng. Data* 65 (9) (2020) 4176–4193, <https://doi.org/10.1021/acs.jced.0c00338>.
- [6] J.M. Calm, The next generation of refrigerants – Historical review, considerations, and outlook, *Int. J. Refrig.* 31 (7) (2008) 1123–1133, <https://doi.org/10.1016/j.jrefrig.2008.01.013>.
- [7] *Montreal protocol on substances that deplete the ozone layer*, 1987.
- [8] *Amendment to the montreal protocol on substances that deplete the ozone layer*, 2016.
- [9] H.P.H. Arp, A. Gredelj, J. Glüge, M. Scheringer, I.T. Cousins, The global threat from the irreversible accumulation of trifluoroacetic acid (TFA), *Environ. Sci. Technol.* 58 (45) (2024) 19925–19935, <https://doi.org/10.1021/acs.est.4c06189>.
- [10] American Innovation and Manufacturing Act of 2020, Pub. L. No. 116-260, § 103, 134 Stat. 1182 (2020), codified at 42 U.S.C. § 7675.
- [11] REGULATION (EU) 2024/573 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 7 February 2024 on fluorinated greenhouse gases, amending Directive (EU) 2019/1937 and repealing Regulation (EU) No 517/2014.
- [12] M.O. McLinden, J.S. Brown, R. Brignoli, A.F. Kazakov, P.A. Domanski, Limited options for low-global-warming-potential refrigerants, *Nat. Commun.* 8 (1) (2017) 14476, <https://doi.org/10.1038/ncomms14476>.
- [13] C.G. Albà, I.I. Alkhatib, L.F. Vega, F. Llorell, Machine learning integration in thermodynamics: Predicting CO₂ mixture saturation properties for sustainable refrigeration applications, *J. CO₂ Util.* 95 (2025) 103072, <https://doi.org/10.1016/j.jcou.2025.103072>.
- [14] C.G. Albà, I.I. Alkhatib, L.F. Vega, F. Llorell, Mapping the flammability space of sustainable refrigerant mixtures through an artificial neural network based on molecular descriptors, *ACS Sustain. Chem. Eng.* 12 (31) (2024) 11561–11577, <https://doi.org/10.1021/acssuschemeng.4c01961>.
- [15] P.G. Polishchuk, T.I. Madzhidov, A. Varnek, Estimation of the size of drug-like chemical space based on GDB-17 data, *J. Comput. Aided Mol. Des.* 27 (8) (2013) 675–679, <https://doi.org/10.1007/s10822-013-9672-4>.
- [16] A. Alakhdar, B. Poczos, N. Washburn, Diffusion models in De Novo drug design, *J. Chem. Inf. Model.* 64 (19) (2024) 7238–7256, <https://doi.org/10.1021/acs.jcim.4c01107>.
- [17] D. Menon, R. Ranganathan, A generative approach to materials discovery, design, and optimization, *ACS Omega* 7 (30) (2022) 25958–25973, <https://doi.org/10.1021/acsomega.2c03264>.
- [18] C.-Y. Chang, Y.-L. Wu, Y.-C. Huang, J.-J. Ke, H.-H.G. Tsai, Accurate and expandable graph deep learning for accelerating predictive identification of high-potential organic photovoltaics, *Chem. Eng. J.* 536 (2026) 175796, <https://doi.org/10.1016/j.ccej.2026.175796>.
- [19] Q. Zhu, S. Feng, Y. Feng, C. Ma, Z. Chu, K. Tang, C. Mao, Y. Guan, S. Gerech, M. Wan, Application of machine learning in the design of biomedical nanomaterials, *Chem. Eng. J.* 530 (2026) 173408, <https://doi.org/10.1016/j.ccej.2026.173408>.
- [20] M. Ruiz-Botella, M. Sales-Pardo, R. Guimerà, A collaborative constrained graph diffusion model for the generation of realistic synthetic molecules, *Nat. Mach. Intell.* (2026) <https://doi.org/10.1038/s42256-026-01229-5>.
- [21] W.G. Chapman, K.E. Gubbins, G. Jackson, M. Radosz, SAFT: Equation-of-state solution model for associating fluids, *Fluid Phase Equilib.* 52 (1989) 31–38, [https://doi.org/10.1016/0378-3812\(89\)80308-5](https://doi.org/10.1016/0378-3812(89)80308-5).
- [22] A. Klamt, F. Eckert, COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids, *Fluid Phase Equilib.* 172 (1) (2000) 43–72, [https://doi.org/10.1016/S0378-3812\(00\)00357-5](https://doi.org/10.1016/S0378-3812(00)00357-5).
- [23] W. Jin, R. Barzilay, T. Jaakkola, Junction tree variational autoencoder for molecular graph generation, 2019, [arXiv:1802.04364](https://arxiv.org/abs/1802.04364).
- [24] N.D. Cao, T. Kipf, Molgan: An implicit generative model for small molecular graphs, 2022, [arXiv:1805.11973](https://arxiv.org/abs/1805.11973).
- [25] R. Mercado, T. Rastemo, E. Lindelöf, G. Klambauer, O. Engkvist, H. Chen, E.J. Bjerrum, Graph networks for molecular design, 2020, <https://doi.org/10.26434/chemrxiv.12843137.v1>, chemRxiv preprint.
- [26] H. Dai, A. Nazi, Y. Li, B. Dai, D. Schuurmans, Scalable deep generative modeling for sparse graphs, 2020, [arXiv:2006.15502](https://arxiv.org/abs/2006.15502).
- [27] R. Liao, Y. Li, Y. Song, S. Wang, C. Nash, W.L. Hamilton, D. Duvenaud, R. Urtasun, R.S. Zemel, Efficient graph generation with graph recurrent attention networks, 2020, [arXiv:1910.00760](https://arxiv.org/abs/1910.00760).
- [28] S. Lee, J. Jo, S.J. Hwang, Exploring chemical space with score-based out-of-distribution generation, 2023, [arXiv:2206.07632](https://arxiv.org/abs/2206.07632).
- [29] K.K. Haefeli, K. Martinkus, N. Perraudin, R. Wattenhofer, Diffusion models for graphs benefit from discrete state spaces, 2023, [arXiv:2210.01549](https://arxiv.org/abs/2210.01549).
- [30] Y. Qin, C. Vignac, P. Frossard, Sparse training of discrete diffusion models for graph generation, 2024, [arXiv:2311.02142](https://arxiv.org/abs/2311.02142).
- [31] Y. Qin, M. Madeira, D. Thanou, P. Frossard, Defog: Discrete flow matching for graph generation, 2025, [arXiv:2410.04263](https://arxiv.org/abs/2410.04263).
- [32] C. Vignac, I. Krawczuk, A. Siraudin, B. Wang, V. Cevher, P. Frossard, Digress: Discrete denoising diffusion for graph generation, 2023, [arXiv:2209.14734](https://arxiv.org/abs/2209.14734).
- [33] H. Huang, L. Sun, B. Du, W. Lv, Conditional diffusion based on discrete graph structures for molecular graph generation, 2023, [arXiv:2301.00427](https://arxiv.org/abs/2301.00427).
- [34] D. Jovell, G. Alonso, P. Gamallo, R. Gonzalez-Olmos, H. Quinteros-Lama, F. Llorell, Combining molecular modelling approaches for a holistic thermophysical characterisation of fluorinated refrigerant blends, *Int. J. Refrig.* 175 (2025) 412–423, <https://doi.org/10.1016/j.jrefrig.2025.03.026>.
- [35] G. Raabe, Molecular simulation studies on refrigerants past – present – future, *Fluid Phase Equilib.* 485 (2019) 190–198, <https://doi.org/10.1016/j.fluid.2018.12.022>.
- [36] N. Wang, M.N. Carlozo, E. Marin-Rimoldi, B.J. Befort, A.W. Dowling, E.J. Maginn, Machine learning-enabled development of accurate force fields for refrigerants, *J. Chem. Theory Comput.* 19 (14) (2023) 4546–4558, <https://doi.org/10.1021/acs.jctc.3c00338>.
- [37] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.* 118 (45) (1996) 11225–11236, <https://doi.org/10.1021/ja9621760>.
- [38] S. Echeverri Restrepo, G.E. Morales-Espejel, Density and viscosity of refrigerant R123 with OPLS force field and atomic SLOD, *Fluid Phase Equilib.* 605 (2026) 114684, <https://doi.org/10.1016/j.fluid.2026.114684>.
- [39] E. Paulechka, A. Kazakov, M. Frenkel, Monte Carlo simulation of vapor–liquid equilibria for perfluoropropane (R-218) and 2, 3, 3, 3-tetrafluoropropene (R-1234yf), *Int. J. Thermophys.* 31 (3) (2010) 462–474, <https://doi.org/10.1007/s10765-010-0764-2>.
- [40] F.J. Blas, L.F. Vega, Thermodynamic behaviour of homonuclear and heteronuclear Lennard-Jones chains with association sites from simulation and theory, *Mol. Phys.* 92 (1) (1997) 135–150, <https://doi.org/10.1080/002689797170707>.
- [41] C.G. Albà, I.I. Alkhatib, F. Llorell, L.F. Vega, Hunting sustainable refrigerants fulfilling technical, environmental, safety and economic requirements, *Renew. Sustain. Energy Rev.* 188 (2023) 113806, <https://doi.org/10.1016/j.rser.2023.113806>.
- [42] C.G. Albà, I.I. Alkhatib, F. Llorell, L.F. Vega, A novel approach for designing efficient and sustainable cooling cycles with low global warming potential refrigerants, *Appl. Therm. Eng.* 246 (2024) 122895, <https://doi.org/10.1016/j.applthermaleng.2024.122895>.
- [43] D. Duque, J.C. Pàmies, L.F. Vega, Interfacial properties of Lennard-Jones chains by direct simulation and density gradient theory, *J. Chem. Phys.* 121 (22) (2004) 11395–11401, <https://doi.org/10.1063/1.1818679>.
- [44] F. Llorell, R.M. Marcos, L.F. Vega, Free-volume theory coupled with soft-saft for viscosity calculations: comparison with molecular simulation and experimental data, *J. Phys. Chem. B* 117 (27) (2013) 8159–8171, <https://doi.org/10.1021/jp401307t>.
- [45] A. Klamt, F. Eckert, W. Arlt, COSMO-RS: An alternative to simulation for calculating thermodynamic properties of liquid mixtures, *Annu. Rev. Chem. Biomol. Eng.* 1 (2010) 101–122, <https://doi.org/10.1146/annurev-chembioeng-073009-100903>.

- [46] I.I.I. Alkhatib, C.G. Albà, A.S. Darwish, F. Llovel, L.F. Vega, Searching for sustainable refrigerants by bridging molecular modeling with machine learning, *Ind. Eng. Chem. Res.* 61 (21) (2022) 7414–7429, <http://dx.doi.org/10.1021/acs.iecr.2c00719>.
- [47] V. Papaioannou, T. Lafitte, C. Avendaño, C.S. Adjiman, G. Jackson, A. Galindo E.A. Müller, Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments, *J. Chem. Phys.* 140 (5) (2014) 054107, <http://dx.doi.org/10.1063/1.4851455>.
- [48] G. Landrum, P. Tosco, B. Kelley, R. Rodriguez, D. Cosgrove, R. Vianello, sriniker, P. Gedeck, G. Jones, NadineSchneider, E. Kawashima, D. Nealschneider, A. Dalke, M. Swain, B. Cole, S. Turk, A. Savelev, tadhurst cdd, A. Vaucher, M. Wójcikowski, I. Take, V.F. Scalfani, R. Walker, K. Ujihara, D. Probst, J. Lehtivarjo, H. Faara, guillaume godin, A. Pahl, J. Monat, RDKit: Open-Source Cheminformatics, Zenodo, 2025, <http://dx.doi.org/10.5281/ZENODO.591637>, <https://www.rdkit.org>.
- [49] R.L. Shank, Thermodynamic properties of 1,1,1,2,2-pentafluoropropane (refrigerant 245), *J. Chem. Eng. Data* 12 (4) (1967) 474–480, <http://dx.doi.org/10.1021/jc60035a004>.
- [50] D.R. Defibaugh, M.R. Moldover, Compressed and saturated liquid densities for 18 halogenated organic compounds, *J. Chem. Eng. Data* 42 (1) (1997) 160–168, <http://dx.doi.org/10.1021/jc960266e>.
- [51] Z.-q. Yang, L.-g. Kou, S. Han, C. Li, Z.-j. Hao, W. Mao, W. Zhang, J. Lu, Vapor-liquid equilibria of 2,3,3,3-tetrafluoropropene (HFO-1234yf) + 1,1,1,2,2-pentafluoropropane (HFC-245cb) system, *Fluid Phase Equilib.* 427 (2016) 390–393, <http://dx.doi.org/10.1016/j.fluid.2016.07.031>.
- [52] L.A. Weber, D.E. Defibaugh, Vapor pressure of 1, 1, 1, 2, 2-pentafluoropropane, *J. Chem. Eng. Data* 41 (4) (1996) 762–764, <http://dx.doi.org/10.1021/jc950312v>.
- [53] P. Linstrom, NIST chemistry WebBook, NIST standard reference database 69, 1997, <http://dx.doi.org/10.18434/T4D303>.
- [54] S. Devotta, A. Chelani, A. Vonsild, Prediction of global warming potentials of refrigerants and related compounds from their molecular structure – An artificial neural network with group contribution method, *Int. J. Refrig.* 131 (2021) 756–765, <http://dx.doi.org/10.1016/j.ijrefrig.2021.08.011>.
- [55] A. Zini, L. Socci, G. Vaccaro, A. Rocchetti, L. Talluri, Working fluid selection for high-temperature heat pumps: A comprehensive evaluation, *Energies* 17 (7) (2024) 1556, <http://dx.doi.org/10.3390/en17071556>.
- [56] Z. Lu, Y. Gong, Y. Yao, C. Luo, W. Ma, Development of a high temperature heat pump system for steam generation using medium-low temperature geothermal water, *Energy Procedia* 158 (2019) 6046–6054, <http://dx.doi.org/10.1016/j.egypro.2019.01.513>.
- [57] S. Anand, A. Gupta, S. Tyagi, Simulation studies of refrigeration cycles: A review, *Renew. Sustain. Energy Rev.* 17 (2013) 260–277, <http://dx.doi.org/10.1016/j.rser.2012.09.021>.
- [58] A.G. Kusne, H. Yu, C. Wu, H. Zhang, J. Hatrick-Simpers, B. DeCost, S. Sarker, C. Oses, C. Toher, S. Curtarolo, A.V. Davydov, R. Agarwal, L.A. Bendersky, M. Li, A. Mehta, I. Takeuchi, On-the-fly closed-loop materials discovery via Bayesian active learning, *Nat. Commun.* 11 (2020) 5966, <http://dx.doi.org/10.1038/s41467-020-19597-w>.
- [59] N.J. Szymanski, B. Rendy, Y. Fei, R.E. Kumar, T. He, D. Milsted, M.J. McDermott, M. Gallant, E.D. Cubuk, A. Merchant, H. Kim, A. Jain, C.J. Bartel, K. Persson, Y. Zeng, G. Ceder, An autonomous laboratory for the accelerated synthesis of inorganic materials, *Nature* 624 (2023) 86–91, <http://dx.doi.org/10.1038/s41586-023-06734-w>.
- [60] H.H. Loeffler, S. Wan, M. Klähn, A.P. Bhati, P.V. Coveney, Optimal molecular design: Generative active learning combining REINVENT with precise binding free energy ranking simulations, *J. Chem. Theory Comput.* 20 (2024) 8308–8328, <http://dx.doi.org/10.1021/acs.jctc.4c00576>.