

FROM KNOWLEDGE TREE TO KNOWLEDGE FOREST: HAR- NESSING CHEMICAL UNDERSTANDING WITH MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

The 2024 Physics and Chemistry Nobel Prizes to machine learning (ML) and artificial intelligence (AI) breakthroughs marked “*Year 1 of AI for Science*”, underscoring their transformative role in physical sciences. Yet data are not the same as understanding—a distinction central to chemistry, which has long relied on concepts such as bond, aromaticity, and reactivity as scaffolds for understanding and explanation. Building on our recent perspectives (ACS Phys. Chem. Au 2024, 4, 135–142; J. Chem. Theory Compt. 2025, DOI: 10.1021/acs.jctc.5c01299), this article explores how ML/AI can become engines of chemical understanding. We introduce a quintet of chemical knowledge—ontology, epistemology, theory, concept, and understanding—and develop the metaphors of the *Knowledge Tree* and *Knowledge Forest* to show how diverse epistemologies interact and recursively enrich one another. Case studies on aromaticity, catalysis, orbital-free density functional theory, and protein folding illustrate how ML features, when interpreted as conceptual roots, yield fruits of understanding. Contrasting *multiscale modeling* with *hierarchical modeling*, we argue that ML enables emergent, concept-driven integration across levels. Cultivating this plural and hierarchical ecosystem may guide theoretical chemistry toward its next breakthroughs, resolving Dirac’s dilemma not by brute force but by forests of concepts that transform data into enduring understanding.

1 INTRODUCTION

The year 2024 will likely be remembered as “*Year 1 of AI for Science*”. In the same week, the Nobel Prize in Physics recognized John J. Hopfield and Geoffrey Hinton for pioneering artificial neural networks [1; 2], while the Nobel Prize in Chemistry honored Demis Hassabis and John Jumper for AlphaFold [3], alongside David Baker for computational protein design [4; 5]. These awards marked a turning point: machine learning (ML) and artificial intelligence (AI) are no longer peripheral accelerators of computation but central epistemic partners in scientific discovery [6; 7]. Yet it is important to note that the chemistry prize was awarded for solving a 50-year challenge in protein structure prediction—mapping one-dimensional amino acid sequences to three-dimensional protein folds [8]—not for cultivating chemical understanding per se [9; 10].

This distinction matters. Numerical data, however accurate, are not the same as understanding. Dirac’s dilemma [11] reminds us that even though the fundamental laws governing chemistry are known, solving the underlying equations—or bypassing them with data-driven shortcuts—does not automatically yield conceptual insight [12]. Chemistry has always thrived on concepts: bonds [13; 14; 15], aromaticity [16; 17; 18], acidity [19; 20; 21], reactivity [22; 23; 24], and countless others that mediate between abstract theory and empirical phenomena [25; 26; 27]. These concepts are not merely linguistic conveniences; they are the scaffolding of explanation, generalization, and transferability [28; 29; 30; 31]. The central question of this perspective, therefore, is how ML and AI can help us not only predict outcomes but also *harness and extend chemical understanding* [32; 33].

To address this question, we must revisit the foundations of chemical knowledge. Chemistry lies between physics and biology: while physics emphasizes fundamental quantities such as potential energy and angular momentum, and biology highlights functionality and selectivity, chemistry focuses on how structural changes govern reactivity. This intermediate position gives the discipline its distinction. To bridge the abstract physical equation of quantum mechanics with practical chemical insight, we rely on theoretical frameworks such as Valence Bond Theory (VBT) [34], Molecular Orbital Theory (MOT) [35], and Den-

055 sity Functional Theory (DFT) [36; 37; 38; 39]. These theories transform the Schrödinger equation into
056 numerical outputs that can be interpreted in chemical terms, giving rise to concepts such as frontier or-
057 bitals [40; 41], chemical bonds [13; 14; 15], aromatic ring currents [42], and catalytic cycles [43]. Such
058 constructions are not merely computational artifacts; they provide explanatory power, predictive capability,
059 and transferability across diverse systems. Thus, chemical understanding depends less on performing large
060 computations than on mastering these conceptual tools and applying them to interpret, explain, and predict.

061 Against this backdrop, the rise of ML and AI presents both a challenge and an opportunity. On the one
062 hand, ML excels at pattern recognition and data generation, often surpassing traditional theories but without
063 offering clear conceptual interpretation. On the other hand, its internal workings—as high-dimensional dy-
064 namical systems characterized by coarse-graining, emergence, and hierarchy—can themselves be examined
065 through the lens of physics. This perspective opens the door to viewing ML not only as a data-generating
066 tool but also as a *concept-discovery engine*, capable of uncovering new descriptors, regularities, and prin-
067 ciples that expand chemical ontology, deepen chemical epistemology, and ultimately advance chemical
068 understanding.

069 This perspective explores that possibility in a structured way. We begin with a philosophical framework
070 that links ontology, epistemology, theory, concept, and understanding. Building on our earlier work [44;
071 45; 46; 47; 48; 49; 50], we then introduce the “*Knowledge Tree*” of traditional theoretical chemistry and
072 contrast it with the “*Knowledge Forest*” that defines the ML/AI era. To ground this transition, we examine
073 the physics of ML models as an epistemological foundation, focusing on coarse-graining, order parameters,
074 emergence, and hierarchy. From there, we propose a general scheme for harnessing chemical understanding
075 with ML and AI, illustrated by case studies on aromaticity [16; 17; 18], catalysis [44; 45], orbital-free
076 density functional theory [51; 52; 53; 54; 55; 56; 57], and protein folding [3; 4; 5]. We next compare
077 multiscale and hierarchical modeling to show how ML enables concept-driven integration across different
078 levels. Finally, in the Outlook, we return to the Knowledge Forest metaphor, considering how cultivating
079 this plural ecosystem may guide us toward its next breakthroughs where data generation and chemical
080 understanding are intertwined, and ML/AI becomes a genuine *concept-discovery engine*.

081 082 2 QUINTET OF CHEMICAL KNOWLEDGE 083

084 Chemical understanding rests on more than calculations or data. We think in terms of entities, rules, and
085 explanations, and these operate at multiple levels. Earlier frameworks often emphasized a triad of ontology
086 (*what*) [58], epistemology (*how*) [59], and understanding (*why*) [60]. While powerful, this *what-how-why*
087 triad is incomplete: it omits the systematic role of theories and the indispensable scaffolding of concepts.
088 To capture the full dynamics of chemical knowledge, we must recognize a quintet of interdependent com-
089 ponents—ontology, epistemology, theory, concept, and understanding—as illustrated in Figure 1a.

090 Ontology addresses *what exists*. In chemistry, this includes atoms, molecules, bonds, functional groups,
091 and electron densities that span the vast chemical space of all possible compounds. With the rise of ML,
092 ontological scope now extends to data structures such as features and embeddings, which serve as carriers
093 of information. Epistemology concerns *how we know*, encompassing experimental tools, computational
094 methods, and representational practices. For much of the twentieth century, the Schrödinger equation of
095 quantum mechanics provided the primary channel from ontology to knowledge. Theory supplies the struc-
096 tured frameworks that make epistemology effective: VBT [34], MOT [35], and DFT [36; 37; 38; 39]
097 each translate quantum principles into computational data. Concepts as portable and heuristic constructs
098 are the essential scaffolding of chemical reasoning. Bonds, aromaticity, and acidity are not mere labels but
099 indispensable constructs that mediate between theory and practice, enabling explanation, generalization,
100 and transferability. Understanding is the culmination of this sequence, where we can *answer why*—why
101 molecules adopt certain structures, why reactions follow one pathway over another, and why broad prin-
102 ciples unify disparate observations.

103 Within the quintet, concepts occupy a privileged role. They are neither reducible to ontology nor fully
104 determined by epistemology but emerge from the interplay between the two. As the base of chemical
105 reasoning, concepts enable both explanation and transferability. A bond, though not directly measurable,
106 distills stability and connectivity from spectra and wavefunctions. Aromaticity, while not tied to a single
107 observable, unifies magnetic, energetic, and structural criteria into a coherent principle. Acidity and basicity
108 link simple experimental measures to theoretical descriptors of electronic structure. In this sense, concepts
109 are the working currency of chemistry: without them, prediction and understanding would be impossible.
Figure 1a highlights this dual framing by placing concepts at the foundation, underscoring their central role.

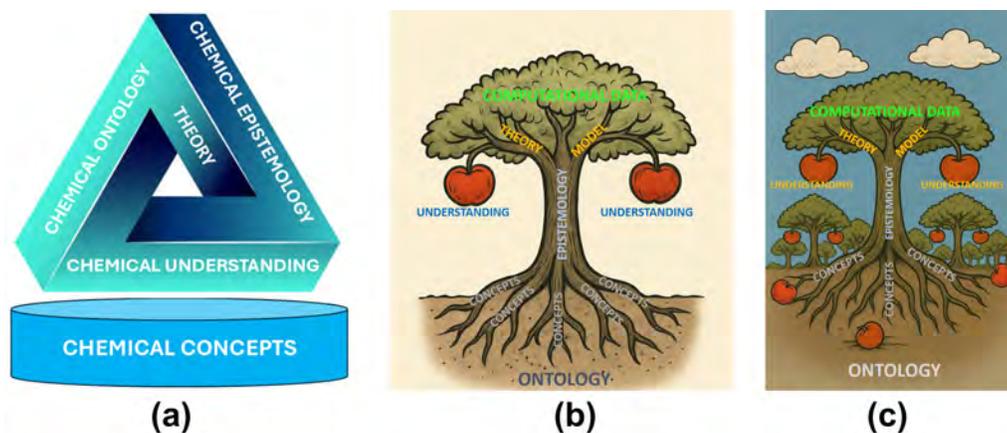


Figure 1: (a) The quintet structure of chemical knowledge. Reproduced with permission from Reference [61]. Copyright 2025, American Chemical Society. (b) The knowledge tree of theoretical chemistry in the 20th century. (c) Knowledge forest of theoretical chemistry in the 21st century. (b) and (c) created by AI.

The quintet is not static but dynamic and recursive. Concepts emerge from ontology yet are reshaped by theory and data. Theories evolve as epistemological frameworks shift. Understanding, in turn, feeds back—refining concepts and inspiring new formulations. This recursive interplay has sustained chemistry for centuries. Aromaticity offers a vivid example: ontology provides π -electrons; epistemology draws on molecular orbital theory; theory codifies Hückel’s $4n + 2$ rule [16]; concepts such as resonance and ring current scaffold reasoning; and understanding arises in recognizing stabilization across benzene, annulenes, and organometallic clusters. Each element of the quintet is indispensable, and their continual circulation fuels the growth of chemical knowledge.

3 KNOWLEDGE TREE: TRADITIONAL THEORETICAL CHEMISTRY

The *knowledge tree* metaphor provides a natural way to capture the intrinsic relationships depicted in Figure 1a and to portray the growth of theoretical chemistry across the past century. As shown in Figure 1b, its roots sink into the soil of ontology—the chemical space itself—drawing nourishment from the fundamental building blocks of electrons, nuclei, atoms, and molecules. From these primitives arise concepts such as orbitals, bonds, resonance, hybridization, aromaticity, acidity, and reactivity, which serve as scaffolds for organizing phenomena. These concepts anchor the tree, stabilizing its growth from the ontological ground of chemical space.

The trunk of the knowledge tree represents epistemology. In twentieth-century theoretical chemistry, this was defined above all by quantum mechanics. The Schrödinger equation provided the central channel linking ontological elements to chemical understanding. In Figure 1b, the trunk rises from concepts into the higher reaches of the tree, symbolizing how quantum mechanics served as the backbone of the field. It transformed the raw soil of ontology into structured knowledge, ensuring that chemical descriptions were not merely heuristic but systematically grounded in first principles.

From this quantum trunk grew the major branches of theoretical chemistry. As shown in Figure 1b, these branches represent theories and models, corresponding to frameworks and their operational implementations. VBT, MOT, and DFT are distinct branches extending from the same epistemological channel, each carrying conceptual insights upward in its own direction. Along these branches sprouted the leaves of computation and, ultimately, the fruits of understanding.

VBT [62; 63; 64] emphasized localized bonding, rooted in atomic orbital overlap, and introduced resonance and hybridization as tools to rationalize molecular stability and geometry. MOT, by contrast, highlighted delocalization, combining atomic orbitals into molecular orbitals that spanned entire molecules. It provided natural explanations for magnetism, conjugation, and spectroscopy, giving rise to models such as Hückel and Hartree–Fock methods [65; 66; 67; 68; 69]. DFT added a third perspective by shifting attention from orbitals to electron density as the fundamental variable. In the Kohn–Sham scheme [36; 37], it blended orbital language with density functionals, creating not only a powerful computational framework but also a family of density-based reactivity descriptors, such as electronegativity, hardness, electrophilicity, and more, that reshaped the conceptual foundations of chemical reactivity theory [70; 71; 72; 73; 74; 75].

165 Together, these branches established the architecture of twentieth-century theoretical chemistry and defined
166 the pathways through which chemical concepts were connected to theory and computation.

167
168 In Figure 1b, the canopy of the tree represents computational data, symbolized by the leaves. As digital
169 computation advanced, theories and models yielded quantitative predictions of energies, structures, and
170 properties. These leaves broadened the tree’s reach by linking theoretical frameworks to experimental ob-
171 servables. Yet the abundance of data brought a familiar challenge: numbers by themselves do not constitute
172 understanding. To become meaningful, they had to be digested back into the conceptual frameworks of
173 chemistry.

174 When this happened, the tree bore fruit. In Figure 1b, these fruits symbolize understanding, the ultimate aim
175 of theoretical chemistry. From VBT came the rationalization of why benzene is more stable than Kekulé
176 structures suggest, or why methane adopts tetrahedral geometry. From MOT came the recognition of oxy-
177 gen’s paramagnetism, the rules governing UV–visible transitions, and the stability patterns of aromatic and
178 antiaromatic systems. From DFT came broad insights into reactivity and catalysis, enabled by descriptors
179 that connected density distributions to chemical behavior. These fruits account for why molecules behave
180 as they do.

181 The knowledge tree of traditional theoretical chemistry was unified in its trunk but diverse in its branches.
182 The Schrödinger equation provided a common epistemological stem, ensuring that all frameworks remained
183 rooted in quantum mechanics. From this foundation, VBT, MOT, and DFT extended outward, each with its
184 own emphasis and style of explanation, yet all part of the same arboreal structure. However, this unity
185 also imposed a constraint: because every branch arose from the same stem, the tree remained mono-
186 lithic. Twentieth-century theoretical chemistry was remarkably fertile, yet its abundant fruit grew from
187 the same tree. Recognizing this limitation sets the stage for the transition to the *knowledge forest*, where
188 new trunks—classical mechanics, statistical mechanics, and machine learning—stand alongside or even
189 mix together with quantum mechanics to cultivate a more plural and interactive ecology of understanding.

190 4 KNOWLEDGE FOREST: THE ML/AI ERA

191 In the twentieth century, theoretical chemistry grew as a single knowledge tree, with quantum mechanics
192 as its trunk and VBT, MOT, and DFT as its main branches. By the twenty-first century, however, this
193 solitary tree has given way to a forest, where multiple epistemological trunks coexist, interact, and enrich
194 one another. The shift is profound: from monism to pluralism, from a single pathway to understanding
195 toward an ecology of interwoven frameworks.

196 Figure 1c illustrates the new metaphor of the knowledge forest. The large tree in the foreground repre-
197 sents the traditional quantum-mechanical trunk, whose branches—VBT, MOT, and DFT—yielded abun-
198 dant leaves of computation and fruits of understanding. In the background, however, other trees have begun
199 to take root in the same ontological soil. These younger but rapidly growing trees symbolize alternative
200 epistemologies—classical mechanics, statistical mechanics, multiscale modeling, machine learning, and
201 quantum computer (not a subject of discussion in this work though)—each developing its own branches,
202 leaves, and fruits. Together, they create a forest where diverse modes of inquiry not only coexist but also
203 interact, enriching one another.

204 The defining feature of the knowledge forest is the presence of multiple trunks, each representing a distinct
205 epistemological channel. Quantum mechanics remains central, but it is no longer alone. Classical mechan-
206 ics, through molecular mechanics and dynamics, captures large-scale conformational behavior. Statistical
207 mechanics grounds thermodynamics, linking microscopic states to macroscopic observables. Multiscale
208 modeling bridges levels of description, as in QM/MM methods [76; 77] that combine enzymic reactions
209 with protein-scale environments. Most recently, machine learning has emerged as a genuinely new trunk,
210 rooted not in physical postulates but in data-driven representation learning. Its ontology is data—structures,
211 densities, spectra, sequences; its epistemology is the learning principle; its “theories” are model architec-
212 tures. ML produces predictions as leaves, but its fruits are new principles, discovered when latent features
213 or embeddings are interpreted as concepts. AlphaFold’s breakthrough in protein structure prediction illus-
214 trates how this trunk can bear a fundamentally different harvest. Together, these trunks grow from the same
215 ontological soil, but channel it differently. The forest metaphor emphasizes that there is no longer a single
216 epistemological authority—theoretical chemistry in the ML/AI era thrives on pluralism.

217 A forest, however, is more than separate trees—it is an ecosystem. Trunks, branches, and canopies overlap,
218 exchanging data and concepts. ML accelerates DFT by suggesting new functionals; statistical mechanics
219 provides ensembles from which ML extracts reduced representations; classical simulations generate data
for ML-based potentials. Conversely, ML enriches multiscale frameworks by coupling across levels of

220 resolution. Such cross-pollination produces hybrid fruits impossible from any single trunk. Understanding
221 in this environment is recursive: insights from one framework enrich and reshape others.

222
223 In this plural setting, chemical concepts evolve and become more dynamic. In the tree era, meanings of
224 bonds, resonance, and aromaticity were relatively stable, each tied to branches of the quantum trunk. In
225 the forest era, ML generates new descriptors from latent variables that capture reactivity or embeddings
226 that reveal structural families. Multiscale modeling creates hierarchical concepts linking molecular events
227 to macroscopic behavior. The fruits of understanding are now a diverse harvest, reflecting plural origins.
228 As Figure 1c illustrates, fruits may also fall to the ground, enriching the soil so that established concepts
229 can be renewed, and new ones can take root. This recursive cycle—where understanding feeds back into
230 concept formation—ensures that the forest does not merely expand but continually regenerates. Yet leaves
231 of data do not automatically ripen into fruits of understanding. Interpretation and conceptual clarity are
232 more crucial than ever. Like a real forest, the knowledge forest must be cultivated if its fruits are to be
nourishing.

233 The transition from a solitary tree to a forest marks a profound shift in the epistemological landscape of
234 theoretical chemistry. Twentieth-century understanding was rooted in a single trunk—quantum mechanics.
235 Today, multiple trunks coexist and interact, producing a richer and more resilient ecosystem. Figure 1c
236 captures this vividly: the quantum tree remains dominant but is now surrounded by others, forming a flourish-
237 ing forest. This pluralism opens new opportunities, but also new responsibilities: predictive power must
238 be balanced by interpretability. The forest must be managed so that its fruits remain fruits of understand-
239 ing, not just numbers. This sets the stage for the next question: how to ground these new epistemologies,
240 especially machine learning, in the physics of models.

241 5 MATHEMATICAL AND PHYSICAL FOUNDATIONS OF ML MODELS 242

243 The knowledge forest of the ML/AI era can flourish only if its new trunks are firmly rooted. In Figure 1c,
244 ML appeared alongside classical mechanics and multiscale modeling, but unlike those physically grounded
245 traditions, it requires special clarification of its epistemological standing. To some, ML appears as an
246 opaque black box. Yet beneath the surface it rests on solid mathematical and physical foundations, and it can
247 be connected to the same epistemological moves that shaped the knowledge tree of traditional theoretical
248 chemistry.

249 *Mathematical foundations.* Four milestones illustrate this grounding. The universal approximation theorem
250 shows that neural networks can approximate any continuous function, granting them a generality compar-
251 able to functional forms long used in chemistry. Geoffrey Hinton’s work on deep learning emphasized hier-
252 archical representation: stacking nonlinear layers allows progressively abstract features to emerge, echoing
253 how physics builds effective theories across scales. Equally essential is the backpropagation algorithm,
254 which applies the chain rule in reverse to compute gradients efficiently, turning learning into a process
255 akin to gradient descent on an energy landscape. Most recently, Kolmogorov–Arnold Networks (KANs)
256 [78; 79; 80] have extended this foundation by replacing scalar weights with learnable univariate functions,
257 often parameterized by splines. Inspired by the Kolmogorov–Arnold representation theorem, KANs offer
258 more compact approximations, improved scaling behavior, and enhanced interpretability, making them at-
259 tractive for scientific applications. A particularly encouraging example is that the latent features that KAN
260 learns from fitting observational data already correspond to physical concepts, e.g., kinetic energy and po-
261 tential energy only from learning to predict the Lagrangian, and the concept of domain wall in learning
262 the geodesics of Schwarzschild black hole [80]. Together, these results provide expressivity, hierarchical
263 structure, trainability, and functional flexibility—positioning ML as a legitimate epistemological trunk, on
264 par with the Schrödinger equation as a generative channel for knowledge.

265 *Features as epistemological roots.* At their core, ML models extract features and latent variables, which can
266 be viewed both as order parameters and as basis sets. As order parameters, they compress complexity into
267 collective descriptors, much like frontier orbitals summarize reactivity or electron density anchors DFT.
268 As basis sets, they expand the representational capacity, just as Gaussian-type functions span molecular
269 orbitals or atomic functions build the total wavefunction. In the tree metaphor, features thus represent epis-
270 temological roots in two ways: channeling nutrients upward by distilling essential patterns and spreading
271 outward by providing scaffolds for approximation. In both roles, they nourish the branches, leaves, and
ultimately the fruits of understanding, making them central to concept formation in the ML era.

272 *Coarse-graining, emergence, and hierarchy.* ML models inherently perform coarse-graining: convolu-
273 tional networks compress local patterns, graph networks pool atomic environments, and transformers cap-
274 ture global interactions. Through these reductions, new variables emerge—much as aromaticity arises

275 from π -electron delocalization—that can serve as conceptual roots and eventually mature into fruits of
276 understanding. This process resembles self-organization, where complex systems spontaneously develop
277 ordered structures, and parallels the renormalization group in physics, which systematically integrates out
278 microscopic details to reveal effective variables at larger scales. It also recalls the ubiquity of scale-free
279 networks, where a few highly connected nodes (hubs) dominate the architecture—much as key features or
280 latent variables in ML can organize and control vast flows of information. Hierarchy is equally central:
281 physics builds successive levels from electrons to molecules to phases, while ML constructs layers from
282 raw embeddings to abstract representations. These analogies suggest that ML does not stand apart from
283 chemistry’s reasoning traditions but echoes their deepest patterns of abstraction.

284 *Duality and representability.* A deeper unity links physical principles and machine-learning formalisms.
285 Many learning dynamics can be expressed as variational processes, governed by loss functionals that play
286 roles analogous to energy minimization in physics: gradient descent mirrors relaxation toward equilibrium;
287 diffusion models parallel stochastic dynamics; and generative networks resemble Hamiltonian flows in
288 latent phase space. Conversely, physical systems can often be recast as learnable mappings, where the
289 governing equations define implicit networks that transform inputs to observables. This duality underscores
290 a shared epistemic logic—both physics and ML acquire knowledge through optimization of a principle.
291 Within this view, the foundational theorems of density functional theory provide a compelling analogy. If a
292 set of learned features suffices to determine the ground-state electron density, then by the Hohenberg–Kohn
293 correspondence those features, collectively, contain all information about the system’s properties. In that
294 sense, ML representations may serve as generalized densities: epistemological sufficient variables from
295 which energies, response functions, and chemical concepts alike can be derived.

296 *Interpretability and concept discovery.* For fruits of understanding to grow, ML outputs (leaves) must be
297 metabolized into conceptual roots through interpretation. Physics reminds us that order parameters matter
298 only when intelligible; similarly, ML features gain significance only when they connect to chemical reason-
299 ing. These features can act both as order parameters, compressing complexity into collective descriptors (as
300 frontier orbitals or electron density do in chemistry), and as basis sets, expanding representational capac-
301 ity much like Gaussian functions span molecular orbitals. When interpreted through tools such as attention
302 maps, latent-space probing, or alignment with chemical descriptors, these features become channels through
303 which raw predictions are transformed into explanatory insight. More recently, development of structured
304 representation learning ignited new opportunities of discovering concepts using ML. Attractive examples
305 are demonstrated by causal representation learning which enables the identification of fundamental con-
306 cepts defining object classes from possibly spuriously correlated data [81; 82], and physics-structured
307 representation learning which composes the concept of Reynolds number in fitting fluid systems [83]. In
308 this way, ML is not merely a predictive tool but a concept discovery engine, capable of generating new
309 descriptors in aromaticity, reactivity, or protein folding. Combined with its mathematical guarantees of
310 universal approximation, its hierarchical architecture, and its efficient learning dynamics, ML is anchored
311 firmly in mathematics and physics. It is therefore not a free-floating black box but a valid epistemological
312 trunk in the knowledge forest—one that connects data back to ontology, cultivates new conceptual roots,
313 and ultimately bears fruits of understanding.

313 6 HARNESSING CHEMICAL UNDERSTANDING WITH ML/AI

314 The preceding sections established ML/AI as a legitimate trunk in the knowledge forest, grounded in math-
315 ematics and physics through expressivity, hierarchy, optimization, and coarse-graining. The natural next
316 step is to ask how this trunk can be cultivated not just for leaves of data generation but for fruits of under-
317 standing. By understanding we mean compact, transferable principles that explain why: why structures are
318 stable, why reactivity follows certain pathways, and why descriptors unify apparently disparate phenomena.

319 The path from data to understanding follows a recursive cycle already familiar from theoretical chemistry:
320 ontological variables such as structures, densities, or spectra are transformed into representations and fea-
321 tures that act as epistemological roots; training with physical constraints and symmetry ensures alignment
322 with chemical reasoning; and interpretation of latent variables and learned descriptors allows data to be me-
323 tabolized into concepts. When these concepts are fed back into practice, they refine the soil of theoretical
324 chemistry, seeding new growth in the forest.

325 In what follows, we illustrate this process through four case studies. Aromaticity shows how ML can re-
326 cover and unify classical descriptors across Hückel, Baird, and Möbius regimes. Catalysis demonstrates
327 how learned features reproduce scaling relations and volcano trends, yielding electronic descriptors that
328 rationalize reactivity and selectivity. Orbital-Free DFT highlights how density-based learning can generate
329 both interpretable functionals and reactivity indices. Protein folding showcases how high-dimensional fea-

330 tures condense into mesoscopic rules and structural motifs [84; 85]. Together, these examples show how
331 ML/AI, properly interpreted, can act as a concept discovery engine—bearing fruits of understanding that
332 enrich the forest of chemical knowledge.

333

334

6.1 AROMATICITY: DISCOVERING UNIFYING DESCRIPTORS

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

6.2 CATALYSIS: SCALING RELATIONS AND CATALYTIC CYCLES

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

Catalysis has long exemplified how data can be distilled into guiding principles. Traditional mechanistic studies, such as those on Mn-based selective catalytic reduction of NO_x , revealed scaling relations—Sabatier volcano trends and linear descriptors—that rationalize activity in terms of adsorption strength and orbital overlap [94; 95]. These descriptors, grounded in ligand field theory and electronic structure, provided conceptual clarity but were necessarily low-dimensional and often oversimplified.

ML extends this framework by uncovering nonlinear, multidimensional features that serve as new epistemological roots [96; 97]. Models trained on large datasets of heterogeneous catalysts generate leaves of data—predicted adsorption energies, activation barriers, or turnover frequencies—that can be probed to reveal emergent variables [98]. Features capturing orbital alignment, spin polarization, or surface-geometry embeddings often correlate with known theories but also extend beyond them. In this sense, ML features act as order parameters, condensing complex reactivity landscapes into a few decisive variables, and as basis sets, expanding the feature space to include higher-order interactions that elude traditional rules [96].

Yet scaling relations and descriptors alone are not sufficient. The catalytic cycle—the network of elementary steps linking reactants, intermediates, and products—is the true conceptual scaffold of catalysis. A cycle captures how activity emerges from the *collective sequence* of transformations, not just isolated events. Studying cycles is essential for design, since the reactivity and selectivity of intermediates, branching of pathways, and identity of turnover-determining steps all govern catalytic performance. From another perspective, catalytic cycles can also be viewed through reaction network theory, where intermediates and pathways form a graph of nodes and edges. This framing emphasizes topology and allows ML to learn not only descriptors but also the network architectures most conducive to efficient turnover, hinting at why certain cycles recur across diverse catalytic systems.

385 Here ML opens a new frontier. By learning across large datasets of catalytic reactions, ML can begin
386 to generate generalized cycle models, identifying recurrent motifs of bond rearrangements, characteristic
387 intermediates, and network topologies that function as emergent conceptual templates. In this vision, ML
388 does not merely reproduce adsorption energies or volcano plots but proposes cycle-level principles that
389 integrate kinetics and thermodynamics. The fruits of understanding are thus rules explaining why certain
390 cycle architectures outperform others, providing a systematic way to predict, compare, and even design
391 catalytic cycles for new systems.

392 The forest metaphor underscores this shift. Just as trees exchange resources through underground networks,
393 catalytic steps interact within cycles, and descriptors cross-pollinate across pathways. ML enriches catalysis
394 by embedding traditional mechanistic principles into a broader epistemology where cycles, descriptors, and
395 emergent features intertwine [96; 97]. In this way, computational screening evolves into conceptual cycle
396 discovery, offering not only descriptors but toward a genuine *theory of catalysis* that can guide the rational
397 design of new and better catalysts [99].

398 6.3 ORBITAL-FREE DFT: LEARNED FUNCTIONALS AND DENSITY-BASED DESCRIPTORS 399

400 DFT epitomizes the knowledge tree: the trunk of quantum mechanics supports a branch where electron den-
401 sity serves as the fundamental variable. Yet its long-standing holy grail—the exact kinetic energy density
402 functional—remains elusive. Orbital-free DFT (OF-DFT) promises linear scaling and broad applicability
403 but has historically been limited by crude approximations. ML offers a way forward by treating kinetic
404 energy as a density functional to be learned directly from data.

405 The universal approximation theorem legitimizes this approach: neural networks, with sufficient depth and
406 parameters, can represent the kinetic energy density functional to arbitrary accuracy. Early studies confirm
407 that ML-learned functionals reproduce Kohn–Sham reference energies, yielding leaves of computational
408 prediction. Sensitivity analyses further highlight emergent density-dependent variables—variance, gradi-
409 ents, local information content—that act as new epistemological roots. These features function both as
410 order parameters, condensing the complexity of many-electron interactions into collective descriptors, and
411 as basis functions, expanding the representational space in which kinetic energy can be faithfully approxi-
412 mated. Recent work on KANs, which replace scalar weights with learnable functional forms, provides an
413 especially promising avenue for constructing interpretable density functionals with improved efficiency and
414 scaling.

415 The true promise of OF-DFT, however, extends beyond efficiency. Because electron density is the ontolog-
416 ical foundation, accurate ML-driven OF-DFT [100; 101; 102; 103; 104; 105; 106; 107] densities would
417 enable a wealth of density-based descriptors central to molecular stability and reactivity. The information-
418 theoretic approach (ITA) [108; 109] is itself an expression of OF-DFT logic. In our earlier work, for exam-
419 ple, the Weizsäcker kinetic energy functional was used to quantify steric effects [110; 111; 112], showing
420 how density functionals can yield quantitative measures of chemical phenomena. Another example is in-
421 formation gain [113] from ITA employed to simultaneously quantify electrophilicity and nucleophilicity
422 [114], the capability of atoms in molecules to donate and accept electrons, respectively. Extending this
423 logic, accurate densities from ML-OF-DFT can generate descriptors for aromaticity, reactivity, and catal-
424 ysis, connecting energy functionals to conceptual chemistry. These applications are particularly valuable
425 for larger and more complex systems—extended materials, biomolecules, catalytic surfaces—where tradi-
426 tional Kohn–Sham methods scale poorly, and hierarchical ML architectures can efficiently capture structure
427 across multiple length scales.

428 In this way, ML contributes more than efficient surrogates: it equips OF-DFT to generate fruits of un-
429 derstanding in the form of quantitative, transferable descriptors. Coarse-grained yet interpretable, these
430 variables provide fresh perspectives on how kinetic energy depends on electron density while simultane-
431 ously empowering density-based frameworks across chemistry. The recursive loop of the forest metaphor is
432 clear: leaves of ML prediction become roots of new concepts, which in turn enrich the soil of density-based
433 theory. OF-DFT thus illustrates how ML can remain firmly anchored in ontology—electron density—while
434 reshaping the conceptual foundations of chemical reactivity theory.

435 6.4 PROTEIN FOLDING: FROM DATA TO PRINCIPLES 436

437 The 2024 Nobel Prize recognized AlphaFold’s triumph in predicting protein structures [115; 116; 117],
438 underscoring ML’s practical impact. Yet the deeper question remains: does AlphaFold contribute to *under-*
439 *standing*? The answer is yes—if one examines how its internal representations function as epistemological
440 roots.

440 Attention maps in AlphaFold highlight correlated residues, effectively serving as order parameters for folding.
441 These maps reveal co-evolutionary couplings and long-range correlations, echoing the physical search
442 for collective variables. The latent space of the model encodes basis-like embeddings that expand sequence
443 information into higher-dimensional representations where structural motifs and contact grammars become
444 separable. Just as functional groups provide transferable building blocks in chemistry, motifs emerge here
445 as conceptual units of protein architecture: recurring patterns that unify sequence diversity under a shared
446 structural principle.

447 This process is inherently hierarchical: folding landscapes emerge through successive levels of organiza-
448 tion—residues to motifs, motifs to domains, domains to folds. The logic parallels hierarchical modeling in
449 physics and chemistry, where coarse-grained variables replace microscopic detail while retaining explana-
450 tory power. Moreover, protein folding can also be viewed as a reaction network, with metastable states as
451 nodes and transitions as edges. From this perspective, AlphaFold does not merely predict endpoints but
452 offers glimpses into the network architecture that makes folding processes robust and reproducible.

453 Here ML acts as a *concept discovery engine*. Instead of relying on hand-crafted descriptors, it uncovers new
454 ones automatically: contact maps, embeddings, and motif grammars that condense the essence of protein
455 structure. The fruits of understanding here are principles about the folding code—constraints, hierarchies,
456 and motifs—that enrich protein science beyond prediction. And as with aromaticity, catalysis, and OF-DFT,
457 these fruits fall back into the soil of concepts, guiding the design of new models and deepening our grasp
458 of biomolecular organization.

459 Taken together, the four case studies reveal a common pattern: ML unifies aromaticity through latent de-
460 localization descriptors, reframes catalysis by uncovering cycle-level principles, empowers OF-DFT with
461 new density-based measures of reactivity, and clarifies protein folding through motifs as transferable con-
462 ceptual units. Each case shows how leaves of data, when interpreted, become conceptual roots that bear
463 fruits of understanding—sustaining the growth of the knowledge forest. This underscores the central bottle-
464 neck: predictive accuracy is no longer the limiting factor, but interpretability and concept formation. Most
465 ML workflows in chemistry remain benchmark-driven, optimized for numbers rather than insight, and as a
466 result many models stop at leaves, never ripening into fruits.

467 The way forward is clear: feature design and selection must become the heart of concept discovery. Hand-
468 crafted features ground ML in ontology—atoms, densities, charges—while learned features open the door
469 to new descriptors such as latent embeddings and attention maps. Balancing the two determines whether
470 ML yields only predictions or genuine principles. When features are deliberately cultivated and connected
471 back to ontology and concepts, ML functions not only as a predictor but as a true engine of chemical
472 understanding, enriching the plural harvest of the knowledge forest. The natural next step is to ask how such
473 conceptual discoveries can be integrated systematically across levels, a challenge that calls for *hierarchical*
474 *modeling*, the subject of the next section.

475 7 FROM MULTISCALE MODELING TO HIERARCHICAL MODELING

476

477 The case studies in the previous section showed that ML can indeed be harnessed for chemical understand-
478 ing. But plural fruits alone are not sufficient. Without structure, they remain scattered insights. The next
479 challenge is to move beyond pluralism toward *hierarchical modeling*, where the diverse contributions of
480 different trunks in the knowledge forest are not only harvested, but also arranged into levels of understand-
481 ing. In this way, predictions and features from ML, once metabolized into concepts, can be organized
482 into a coherent hierarchy that mirrors the layered structure of molecular systems themselves—from elec-
483 trons to bonds, from local descriptors to global reactivity patterns, from specific mechanisms to emergent
484 phenomena.

485 7.1 MULTISCALE VERSUS HIERARCHICAL

486

487 It is important to distinguish multiscale modeling from hierarchical modeling. Multiscale approaches cou-
488 ple different descriptions at different levels of resolution—quantum mechanics for active sites, classical
489 mechanics for environments, continuum models for solvents. Such frameworks are pragmatic, but they
490 often lack conceptual integration. Each scale is treated as a separate module, stitched together by embed-
491 ding schemes or boundary conditions. Figure 2 illustrates this contrast. On the left, multiscale modeling is
492 represented as a patchwork: separate scales coexist but remain siloed, with communication managed only
493 through couplings at their boundaries. On the right, hierarchical modeling is shown as a nested structure:
494 higher levels emerge from lower ones while retaining interpretability in their own right. Just as molecular
orbitals give rise to functional groups, which in turn give rise to reactivity classes, hierarchical modeling

builds a scaffold where concepts at each level are grounded in physics yet stand as distinct explanatory layers. This conceptual nesting transforms mere coexistence into genuine integration.

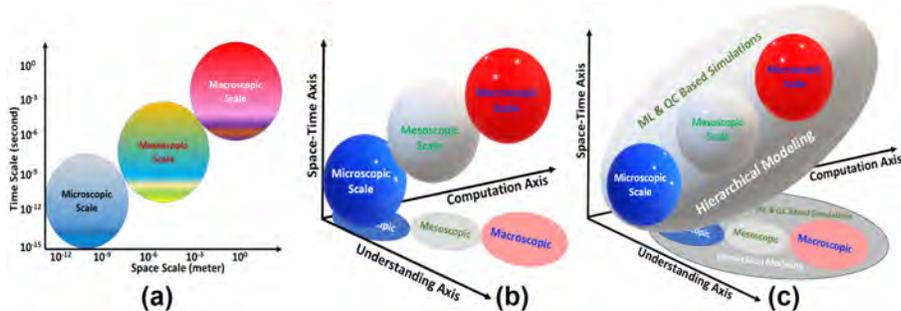


Figure 2: (a) Multiscale modeling along the space and time domains; (b) an understanding axis is added in the multiscale modeling; and (c) schematic representation of hierarchical modeling spanning different hierarchies using ML and QC based simulations. Reproduced with permission from Reference [61]. Copyright 2025, American Chemical Society.

7.2 HIERARCHIES IN PHYSICS, CHEMISTRY, AND ML

Chemistry has always been hierarchical: atoms \rightarrow molecules \rightarrow materials \rightarrow phases \rightarrow macroscopic properties. At each level, concepts—bond, orbital, lattice, state function—cannot be reduced to the lower level, yet they remain connected. The knowledge tree of Section 3 captured this nesting within the quantum trunk, while the knowledge forest of Section 4 broadened it to plural epistemologies. Section 5 showed that ML itself builds hierarchies internally, from raw inputs to latent embeddings to final outputs. Section 6 demonstrated how such learned hierarchies map onto chemical concepts: aromaticity distilled into delocalization descriptors, catalysis organized into cycle-level rules, OF-DFT enriched with density-based measures, and protein folding clarified by motifs.

The opportunity now is to align these parallel hierarchies—those of chemistry, physics, and ML—into a single epistemological framework [118]. Figure 2 provides a visual guide: what was once modular and fragmented under multiscale modeling schemes can now be reframed as a hierarchy, where concepts at one level emerge from and enrich those at another. In this sense, hierarchical modeling is the epistemological culmination of the forest metaphor: different trunks contribute fruits, but hierarchy arranges them into an ecosystem of levels, recursively enriched through feedback.

7.3 PRACTICAL PATHWAYS TO HIERARCHICAL MODELING

How can this vision be realized? Several pathways are emerging. *Representation learning* across levels allows ML to embed atoms into molecules, molecules into reaction networks, and reactions into broader landscapes—making intermediate concepts explicit. *Bridging physics-based and ML hierarchies* turns correlations into organized nesting: in OF-DFT, ML-learned functionals sit between raw density and total energy; in catalysis, descriptors link local orbital alignment to macroscopic activity. *Recursive enrichment* ensures feedback: AlphaFold’s attention maps suggest order parameters that guide coarse-grained folding models, which then refine residue-level features. Finally, *concept bottlenecks*—architectural layers designed to map onto human-interpretable quantities such as aromatic stabilization, bond polarity, or hardness—embed hierarchy directly into model structure.

7.4 THE PROMISE OF HIERARCHICAL UNDERSTANDING

The promise of hierarchical modeling is to move chemistry beyond multiscale coexistence toward conceptual nesting. Instead of merely coupling separate models, we cultivate frameworks where concepts are recursively enriched across levels. Aromaticity becomes not just a rule for electron counts but part of a hierarchy linking orbital structure, energetic stabilization, and reactivity. Protein folding becomes not just a prediction task, but a hierarchy of constraints, motifs, and architectures grounded in sequence. Catalysis becomes not just volcano plots but a hierarchy connecting orbital alignment, scaling relations, cycles, and design rules.

Hierarchical modeling thus represents the next frontier of the knowledge forest. It promises to unify the plural fruits of ML and traditional theory into an organized ecosystem, where concepts at each level are both interpretable and connected. Figure 2 crystallizes this vision: multiscale modeling offered pragmatic

550 couplings, but hierarchical modeling offers conceptual integration. By embedding hierarchy systemati-
551 cally, we can cultivate not just predictive accuracy but genuine understanding—the kind that explains *why*
552 phenomena emerge and how principles recur across chemistry’s many scales.

554 8 OUTLOOK: CULTIVATING THE KNOWLEDGE FOREST TOWARDS NEXT 555 BREAKTHROUGHS 556

557 The journey from the solitary knowledge tree of twentieth-century chemistry to today’s plural knowledge
558 forest marks a shift from conceptual monism to pluralism. New trunks—classical mechanics, statistical me-
559 chanics, multiscale modeling, machine learning, and quantum computing—now grow alongside quantum
560 mechanics. The challenge ahead is not simply to harvest fruits from each trunk, but to cultivate the forest as
561 a whole. The tree of the past gave stability; the forest of the future offers resilience and richness. Cultivating
562 this forest—through pluralism, interpretability, and hierarchy—will ensure that ML/AI contributes not only
563 to large-scale prediction, but to understanding that endures.

564 Three priorities stand out. First, conceptual pluralism must be embraced as a strength. Aromaticity, catal-
565 ysis, orbital-free DFT, and protein folding all show that different epistemologies can coexist, each offering
566 complementary insights. Second, interpretability and emergence must be treated as central design goals.
567 Leaves of computational data become fruits of chemical understanding only when features and latent vari-
568 ables are interpreted as roots that connect back to ontology and concepts; but equally, we must recognize
569 that new conceptual layers emerge from this process, carrying explanatory power that cannot be reduced to
570 lower levels. Feature design and selection are therefore not mere technicalities, but opportunities for both
571 concept discovery and the cultivation of emergent principles. Third, hierarchical modeling is the natural
572 next step. Rather than linking scales pragmatically, we must embed conceptual nesting so that insights at
573 one level enrich those above and below, with fruits falling back to nourish the soil of future concepts.

574 History shows that breakthroughs in theoretical and computational chemistry have been recognized by the
575 Nobel Prize in Chemistry roughly every decade and a half: chemical reactivity theory in 1981, electron
576 transfer theory in 1992, density functional theory in 1998, multiscale modeling in 2013, and protein struc-
577 ture prediction by ML in 2024. If this rhythm continues, by around 2040 and 2055 one can imagine Nobel
578 Prizes awarded for *hierarchical modeling* that unifies physics-based and ML-derived descriptors into coher-
579 ent levels of explanation, or for *concept discovery engines* that propose new principles redefining bonding,
580 reactivity, catalysis, and materials or drug design. Such achievements would mark the full maturation of
581 the knowledge forest—an ecosystem where plural epistemologies and hierarchical modeling together yield
582 an abundant and enduring harvest of chemical understanding. In this way, the dilemma that Dirac posed
583 nearly a century ago—that the equations of physics are known but too difficult to solve—may finally find
584 resolution, not by brute force, but by cultivating forests of concepts that transform data generation into
585 genuine chemical understanding.

586 REFERENCES 587

- 588 [1] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abili-
589 ties. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. doi: 10.1073/pnas.
590 79.8.2554.
- 591 [2] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief
592 nets. *Neural Computation*, 18(7):1527–1554, 2006. doi: 10.1162/neco.2006.18.7.1527.
- 593 [3] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger,
594 Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens
595 Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav
596 Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy,
597 Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Boden-
598 stein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and
599 Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):
600 583–589, 2021. doi: 10.1038/s41586-021-03819-2.
- 601 [4] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie
602 Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park,
603 Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alber-
604 dina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea

- 605 Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christo-
606 pher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate predic-
607 tion of protein structures and interactions using a three-track neural network. *Science*, 373(6557):
608 871–876, 2021. doi: 10.1126/science.abj8754.
- 609 [5] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky,
610 A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan,
611 B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep
612 learning-based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022. doi:
613 10.1126/science.add2187.
- 614 [6] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak,
615 Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne Bergen, Carla P.
616 Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun Manrai,
617 Debora Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Veličković, Max Welling,
618 Linfeng Zhang, Connor W. Coley, Yoshua Bengio, and Marinka Zitnik. Scientific discovery in the
619 age of artificial intelligence. *Nature*, 620(7972):47–60, 2023. doi: 10.1038/s41586-023-06221-2.
- 620 [7] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie
621 Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of*
622 *Modern Physics*, 91(4), 2019. doi: 10.1103/revmodphys.91.045002.
- 623 [8] Ken A. Dill and Justin L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338
624 (6110):1042–1046, 2012. doi: 10.1126/science.1219021.
- 625 [9] Keith T. Butler, Daniel W. Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine
626 learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018. doi: 10.1038/
627 s41586-018-0337-2.
- 628 [10] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer. Unifying machine
629 learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature*
630 *Communications*, 10(1), 2019. doi: 10.1038/s41467-019-12875-2.
- 631 [11] Ana Simões. Dirac’s claim and the chemists. *Physics in Perspective*, 4(3):253–266, 2002. doi:
632 10.1007/s00016-002-8369-1.
- 633 [12] Yun-Fei Shi, Zheng-Xin Yang, Sicong Ma, Pei-Lin Kang, Cheng Shang, P. Hu, and Zhi-Pan Liu.
634 Machine learning for chemistry: Basics and applications. *Engineering*, 27:70–83, 2023. doi: 10.
635 1016/j.eng.2023.04.013.
- 636 [13] Gilbert N. Lewis. The atom and the molecule. *Journal of the American Chemical Society*, 38(4):
637 762–785, 1916. doi: 10.1021/ja02261a002.
- 638 [14] L. Pauling. *The Nature of the Chemical Bond*, volume 18. Cornell university press, 1960.
- 639 [15] R. F. W. Bader. *Atoms in molecules: a quantum theory*. Oxford University Press, Oxford, 1990.
- 640 [16] Erich Hückel. Quantentheoretische beiträge zum benzolproblem. *Zeitschrift für Physik*, 70(3–4):
641 204–286, 1931. doi: 10.1007/BF01339530.
- 642 [17] Tadeusz Marek Krygowski and Michał Ksawery Cyrański. Structural aspects of aromaticity. *Chem-*
643 *ical Reviews*, 101(5):1385–1420, 2001. doi: 10.1021/cr990326u.
- 644 [18] Ferran Feixas, Eduard Matito, Jordi Poater, and Miquel Solà. Quantifying aromaticity with elec-
645 tron delocalisation measures. *Chemical Society Reviews*, 44(18):6434–6451, 2015. doi: 10.1039/
646 c5cs00066a.
- 647 [19] JN Brønsted. Some remarks on the concept of acids and bases. In *Source Book in Chemistry*,
648 1900–1950, volume 42, pages 718–728. Harvard University Press, 1923. doi: 10.4159/harvard.
649 9780674366701.c69.
- 650 [20] G. N. Lewis. *Valence and the Structure of Atoms and Molecules*. Number 14. Chemical Catalog
651 Company, Incorporated, 1923.

- 660 [21] Ralph G. Pearson. Hard and soft acids and bases. *Journal of the American Chemical Society*, 85
661 (22):3533–3539, 1963. doi: 10.1021/ja00905a001.
- 662 [22] R. G. Parr and Yang W. *Density-Functional Theory of Atoms and Molecules*. Density-Functional
663 Theory of Atoms and Molecules, 1995. doi: 10.1093/oso/9780195092769.001.0001.
- 664 [23] I. Fleming. *Molecular Orbitals and Organic Chemical Reactions*. John Wiley & Sons, Chichester,
665 2009.
- 666 [24] Shubin Liu. *Conceptual Density Functional Theory: Towards a New Chemical Reactivity Theory*.
667 Wiley-VCH GmbH, Weinheim, 1st edition, 2022. doi: 10.1002/9783527837941.
- 668 [25] R. Hoffmann. *The same and not the same*. Columbia University Press, 1995.
- 669 [26] J Nye Mary. From chemical philosophy to theoretical chemistry, 1993.
- 670 [27] Mauro Causá, Andreas Savin, and Bernard Silvi. Atoms and bonds in molecules and chemical
671 explanations. *Foundations of Chemistry*, 16(1):3–26, 2013. doi: 10.1007/s10698-013-9192-2.
- 672 [28] M. Johnson and G. Lakoff. *Metaphors we live by*, volume 1. University of Chicago press Chicago,
673 1980.
- 674 [29] Ingo Brigandt. The epistemic goal of a concept: accounting for the rationality of semantic change
675 and variation. *Synthese*, 177(1):19–40, 2009. doi: 10.1007/s11229-009-9623-8.
- 676 [30] Dongbo Zhao, Yilin Zhao, Enhua Xu, Wenqi Liu, Paul W. Ayers, Shubin Liu, and Dahua Chen.
677 Fragment-based deep learning for simultaneous prediction of polarizabilities and nmr shieldings
678 of macromolecules and their aggregates. *Journal of Chemical Theory and Computation*, 20(6):
679 2655–2665, 2024. doi: 10.1021/acs.jctc.3c01415.
- 680 [31] John A. Keith, Valentin Vassilev-Galindo, Bingqing Cheng, Stefan Chmiela, Michael Gastegger,
681 Klaus-Robert Müller, and Alexandre Tkatchenko. Combining machine learning and computational
682 chemistry for predictive insights into chemical systems. *Chemical Reviews*, 121(16):9816–9872,
683 2021. doi: 10.1021/acs.chemrev.1c00107.
- 684 [32] O. Anatole von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Exploring chemical
685 compound space with quantum-based machine learning. *Nature Reviews Chemistry*, 4(7):347–358,
686 2020. doi: 10.1038/s41570-020-0189-9.
- 687 [33] Frank Noé, Alexandre Tkatchenko, Klaus-Robert Müller, and Cecilia Clementi. Machine learning
688 for molecular simulation. *Annual Review of Physical Chemistry*, 71(1):361–390, 2020. doi: 10.
689 1146/annurev-physchem-042018-052331.
- 690 [34] Chen Zhou, Xun Wu, Fuming Ying, and Wei Wu. Ab initio valence bond theory for strongly cor-
691 related systems. *Journal of Chemical Theory and Computation*, 21(13):6243–6257, 2025. doi:
692 10.1021/acs.jctc.5c00596.
- 693 [35] M. J. S. Dewar. A molecular orbital theory of organic chemistry. i. general principles. *Journal of the*
694 *American Chemical Society*, 74(13):3341–3345, 1952. doi: 10.1021/ja01133a038.
- 695 [36] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871,
696 1964. doi: 10.1103/physrev.136.b864.
- 697 [37] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects.
698 *Physical Review*, 140(4A):A1133–A1138, 1965. doi: 10.1103/physrev.140.a1133.
- 699 [38] John P. Perdew. Jacob’s ladder of density functional approximations for the exchange-correlation
700 energy. In *AIP Conference Proceedings*, volume 577, page 1–20, 2001. doi: 10.1063/1.1390175.
- 701 [39] Eberhard Engel and Reiner M. Dreizler. *Density Functional Theory: An Advanced Course*. Springer
702 Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-14090-7.
- 703 [40] Kenichi Fukui, Teijiro Yonezawa, and Haruo Shingu. A molecular orbital theory of reactivity in
704 aromatic hydrocarbons. *The Journal of Chemical Physics*, 20(4):722–725, 1952. doi: 10.1063/1.
705 1700523.

- 715 [41] Kenichi Fukui. Role of frontier orbitals in chemical reactions. *Science*, 218(4574):747–754, 1982.
716 doi: 10.1126/science.218.4574.747.
- 717 [42] Dage Sundholm, Maria Dimitrova, and Raphael J. F. Berger. Current density and molecular magnetic
718 properties. *Chemical Communications*, 57(93):12362–12378, 2021. doi: 10.1039/d1cc03350f.
- 719 [43] Sebastian Kozuch and Sason Shaik. How to conceptualize catalytic cycles? the energetic span model.
720 *Accounts of Chemical Research*, 44(2):101–110, 2010. doi: 10.1021/ar1000956.
- 721 [44] Shubin Liu. *Exploring Chemical Concepts Through Theory and Computation*. Wiley, 2024. doi:
722 10.1002/9783527843435.
- 723 [45] Shubin Liu and Xiaojuan Zhang. Chemical concepts from density functional theory. *Acta Physico-*
724 *Chimica Sinica*, 34(6):563–566, 2018. doi: 10.3866/PKU.WHXB201802282.
- 725 [46] Shubin Liu. Harvesting chemical understanding with machine learning and quantum computers.
726 *ACS Physical Chemistry Au*, 4(2):135–142, 2024. doi: 10.1021/acspchemau.3c00067.
- 727 [47] Shubin Liu. Homochirality originates from the handedness of helices. *The Journal of Physical*
728 *Chemistry Letters*, 11(20):8690–8696, 2020. doi: 10.1021/acs.jpcclett.0c02144.
- 729 [48] Shubin Liu, Chuying Rong, and Tian Lu. Information conservation principle determines elec-
730 trophilicity, nucleophilicity, and regioselectivity. *The Journal of Physical Chemistry A*, 118(20):
731 3698–3704, 2014. doi: 10.1021/jp5032702.
- 732 [49] Shubin Liu. Quantifying reactivity for electrophilic aromatic substitution reactions with hirshfeld
733 charge. *The Journal of Physical Chemistry A*, 119(12):3107–3111, 2015. doi: 10.1021/acs.jpca.
734 5b00443.
- 735 [50] Shubin Liu. Where does the electron go? the nature of ortho/para and meta group directing in
736 electrophilic aromatic substitution. *The Journal of Chemical Physics*, 141(19), 2014. doi: 10.1063/
737 1.4901898.
- 738 [51] Yan Alexander Wang and Emily A. Carter. *Orbital-Free Kinetic-Energy Density Functional Theory*,
739 page 117–184. Kluwer Academic Publishers, 2002. doi: 10.1007/0-306-46949-9_5.
- 740 [52] Wenhui Mi, Kai Luo, S. B. Trickey, and Michele Pavanello. Orbital-free density functional the-
741 ory: An attractive electronic structure method for large-scale first-principles simulations. *Chemical*
742 *Reviews*, 123(21):12039–12104, 2023. doi: 10.1021/acs.chemrev.2c00758.
- 743 [53] Junchao Xia, Chen Huang, Ilgyou Shin, and Emily A. Carter. Can orbital-free density functional the-
744 ory simulate molecules? *The Journal of Chemical Physics*, 136(8), 2012. doi: 10.1063/1.3685604.
- 745 [54] Wenhui Mi, Xuecheng Shao, Chuanxun Su, Yuanyuan Zhou, Shoutao Zhang, Quan Li, Hui Wang,
746 Lijun Zhang, Maosheng Miao, Yanchao Wang, and Yanming Ma. Atlas: A real-space finite-
747 difference implementation of orbital-free density functional theory. *Computer Physics Communi-*
748 *cations*, 200:87–95, 2016. doi: 10.1016/j.cpc.2015.11.004.
- 749 [55] Matthew S. Ryley, Michael Withnall, Tom J. P. Irons, Trygve Helgaker, and Andrew M. Teale.
750 Robust all-electron optimization in orbital-free density-functional theory using the trust-region im-
751 age method. *The Journal of Physical Chemistry A*, 125(1):459–475, 2021. doi: 10.1021/acs.jpca.
752 0c09502.
- 753 [56] Valentin V Karasiev, Katerina P Hilleke, and S B Trickey. Free-energy orbital-free density functional
754 theory: recent developments, perspective, and outlook. *Electronic Structure*, 7(1):013001, 2025. doi:
755 10.1088/2516-1075/adadd4.
- 756 [57] Vincent L Lignères and Emily A Carter. An introduction to orbital-free density functional the-
757 ory. In *Handbook of materials modeling: methods*, pages 137–148. Springer, 2005. doi:
758 10.1007/978-1-4020-3286-8_9.
- 759 [58] Claudia Bobach, Timo Böhme, Ulf Laube, Anett Püschel, and Lutz Weber. Automated compound
760 classification using a chemical ontology. *Journal of Cheminformatics*, 4(1), 2012. doi: 10.1186/
761 1758-2946-4-40.

- 770 [59] Hrvoj Vančik. Chemistry and philosophy of science. In *Philosophy of Chemistry*, pages 1–18.
771 Springer, 2021. doi: 10.1007/978-3-030-69107-7_1.
772
- 773 [60] Vicente Talanquer. Chemical rationales: another triplet for chemical thinking. *International Journal*
774 *of Science Education*, 40(15):1874–1890, 2018. doi: 10.1080/09500693.2018.1513671.
775
- 776 [61] Shubin Liu. The evolving quest for chemical understanding in the quantum age. *Journal of Chemical*
777 *Theory and Computation*, 2025. doi: 10.1021/acs.jctc.5c01299.
- 778 [62] Lingchun Song, Yirong Mo, Qianer Zhang, and Wei Wu. Xmvb: A program for ab initio nonorthog-
779 onal valence bond computations. *Journal of Computational Chemistry*, 26(5):514–521, 2005. doi:
780 10.1002/jcc.20187.
- 781 [63] Peifeng Su, Lingchun Song, Wei Wu, Philippe C. Hiberty, and Sason Shaik. Valence bond calcula-
782 tions of hydrogen transfer reactions: A general predictive pattern derived from theory. *Journal of the*
783 *American Chemical Society*, 126(41):13539–13549, 2004. doi: 10.1021/ja048105f.
784
- 785 [64] Yirong Mo, Jiali Gao, and Sigrid D. Peyerimhoff. Energy decomposition analysis of intermolecular
786 interactions using a block-localized wave function approach. *The Journal of Chemical Physics*, 112
787 (13):5530–5538, 2000. doi: 10.1063/1.481185.
- 788 [65] J. C. Slater. The theory of complex spectra. *Physical Review*, 34(10):1293–1322, 1929. doi: 10.
789 1103/physrev.34.1293.
790
- 791 [66] D. R. Hartree. Some relations between the optical spectra of different atoms of the same electronic
792 structure. ii. aluminium-like and copper-like atoms. *Mathematical Proceedings of the Cambridge*
793 *Philosophical Society*, 23(3):304–326, 1926.
794
- 795 [67] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift*
796 *für Physik*, 61(1–2):126–148, 1930. doi: 10.1007/bf01340294.
- 797 [68] C. C. J. Roothaan. New developments in molecular orbital theory. *Reviews of Modern Physics*, 23
798 (2):69–89, 1951. doi: 10.1103/revmodphys.23.69.
799
- 800 [69] Toshikatsu Koga, Hiroshi Tatewaki, and Ajit J. Thakkar. Roothaan-hartree-fock wave functions for
801 atoms withz54. *Physical Review A*, 47(5):4510–4512, 1993. doi: 10.1103/physreva.47.4510.
- 802 [70] Shubin Liu, Chunying Rong, and Tian Lu. Electronic forces as descriptors of nucleophilic and
803 electrophilic regioselectivity and stereoselectivity. *Physical Chemistry Chemical Physics*, 19(2):
804 1496–1503, 2017. doi: 10.1039/c6cp06376d.
- 805 [71] Yilin Zhao, Dongbo Zhao, Shubin Liu, Chunying Rong, and Paul W. Ayers. Why are information-
806 theoretic descriptors powerful predictors of atomic and molecular polarizabilities. *Journal of Molec-*
807 *ular Modeling*, 30(11), 2024. doi: 10.1007/s00894-024-06162-1.
808
- 809 [72] Shu-Bin LIU. Information-theoretic approach in density functional reactivity theory. *Acta Physico-*
810 *Chimica Sinica*, 32(1):98–118, 2016. doi: 10.3866/pku.whxb201510302.
811
- 812 [73] Xiaofang Cao, Chunying Rong, Aiguo Zhong, Tian Lu, and Shubin Liu. Molecular acidity: An accu-
813 rate description with information-theoretic approach in density functional reactivity theory. *Journal*
814 *of Computational Chemistry*, 39(2):117–129, 2017. doi: 10.1002/jcc.25090.
- 815 [74] Shubin Liu, Cynthia K. Schauer, and Lee G. Pedersen. Molecular acidity: A quantitative conceptual
816 density functional theory description. *The Journal of Chemical Physics*, 131(16), 2009. doi: 10.
817 1063/1.3251124.
818
- 819 [75] Shubin Liu and Lee G. Pedersen. Estimation of molecular acidity via electrostatic potential at the nu-
820 cleus and valence natural atomic orbitals. *The Journal of Physical Chemistry A*, 113(15):3648–3655,
821 2009. doi: 10.1021/jp811250r.
- 822 [76] Meiyi Liu, Yingjie Wang, Yakun Chen, Martin J. Field, and Jiali Gao. Qm/mm through the 1990s:
823 The first twenty years of method development and applications. *Israel Journal of Chemistry*, 54
824 (8–9):1250–1263, 2014. doi: 10.1002/ijch.201400036.

- 825 [77] Jiali Gao, John Z. H. Zhang, and Kendall N. Houk. Beyond qm/mm: Fragment quantum mechanical
826 methods. *Accounts of Chemical Research*, 47(9):2711–2711, 2014. doi: 10.1021/ar500293u.
827
- 828 [78] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić,
829 Thomas Y. Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks, 2025.
- 830 [79] A. N. Kolmogorov. Publications of a. n. kolmogorov. *The Annals of Probability*, 17(3):945–964,
831 1989.
832
- 833 [80] Ziming Liu, Pingchuan Ma, Yixuan Wang, Wojciech Matusik, and Max Tegmark. Kan 2.0:
834 Kolmogorov-arnold networks meet science, 2024.
- 835 [81] Chang Liu, Xinwei Sun, Jindong Wang, Haoyue Tang, Tao Li, Tao Qin, Wei Chen, and Tie-Yan
836 Liu. Learning causal semantic representation for out-of-distribution prediction. In M. Ranzato,
837 A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural
838 Information Processing Systems*, volume 34, pages 6155–6170. Curran Associates, Inc., 2021.
839
- 840 [82] Xinwei Sun, Botong Wu, Xiangyu Zheng, Chang Liu, Wei Chen, Tao Qin, and Tie-Yan Liu. Recov-
841 ering latent causal factor for generalization to distributional shifts. In M. Ranzato, A. Beygelzimer,
842 Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Process-
843 ing Systems*, volume 34, pages 16846–16859. Curran Associates, Inc., 2021.
- 844 [83] Tingxiong Xiao, Xinxin Song, Ziqian Wang, Boyang Zhang, and Jinli Suo. A unified data-driven
845 framework for efficient scientific discovery, 2025.
846
- 847 [84] Ka-Chun Wong. Dna motif recognition modeling from protein sequences. *iScience*, 7:198–211,
848 2018. doi: 10.1016/j.isci.2018.09.003.
- 849 [85] J. Liu. Combining selex with quantitative assays to rapidly obtain accurate models of protein-dna
850 interactions. *Nucleic Acids Research*, 33(17):e141–e141, 2005. doi: 10.1093/nar/gni139.
851
- 852 [86] Paul von Ragué Schleyer, Christoph Maerker, Alk Dransfeld, Haijun Jiao, and Nicolaas J. R. van
853 Eikema Hommes. Nucleus-independent chemical shifts: A simple and efficient aromaticity probe.
854 *Journal of the American Chemical Society*, 118(26):6317–6318, 1996. doi: 10.1021/ja960582d.
855
- 856 [87] Zhongfang Chen, Chaitanya S. Wannere, Clémence Corminboeuf, Ralph Puchta, and Paul von Ragué
857 Schleyer. Nucleus-independent chemical shifts (nics) as an aromaticity criterion. *Chemical Reviews*,
858 105(10):3842–3888, 2005. doi: 10.1021/cr030088+.
- 859 [88] Eduard Matito, Miquel Duran, and Miquel Solà. The aromatic fluctuation index (flu): A new aro-
860 maticity index based on electron delocalization. *The Journal of Chemical Physics*, 122(1), 2004.
861 doi: 10.1063/1.1824895.
- 862 [89] J. Kruszewski and T.M. Krygowski. Definition of aromaticity basing on the harmonic oscillator
863 model. *Tetrahedron Letters*, 13(36):3839–3842, 1972. doi: 10.1016/s0040-4039(01)94175-9.
864
- 865 [90] Tadeusz Marek Krygowski. Crystallographic studies of inter- and intramolecular interactions re-
866 flected in aromatic character of .pi.-electron systems. *Journal of Chemical Information and Com-
867 puter Sciences*, 33(1):70–78, 1993. doi: 10.1021/ci00011a011.
- 868 [91] Siamak Noorizadeh and Ehsan Shakerzadeh. Shannon entropy as a new measure of aromaticity,
869 shannon aromaticity. *Physical Chemistry Chemical Physics*, 12(18):4742–4749, 2010. doi: 10.
870 1039/b916509f.
871
- 872 [92] Meng Li, Xinjie Wan, Chunying Rong, Dongbo Zhao, and Shubin Liu. Directionality and ad-
873 ditivity effects of molecular acidity and aromaticity for substituted benzoic acids under exter-
874 nal electric fields. *Physical Chemistry Chemical Physics*, 25(40):27805–27816, 2023. doi:
875 10.1039/d3cp02982d.
- 876 [93] Wenjie Wu, Zemin Wu, Chunying Rong, Tian Lu, Ying Huang, and Shubin Liu. Computational
877 study of chemical reactivity using information-theoretic quantities from density functional reactivity
878 theory for electrophilic aromatic substitution reactions. *The Journal of Physical Chemistry A*, 119
879 (29):8216–8224, 2015. doi: 10.1021/acs.jpca.5b04309.

- 880 [94] A. A. Balandin. *Modern State of the Multiplet Theor of Heterogeneous Catalysis*, page 1–210.
881 Elsevier, 1969. doi: 10.1016/s0360-0564(08)60029-2.
- 882 [95] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen. Towards the computational design
883 of solid catalysts. *Nature Chemistry*, 1(1):37–46, 2009. doi: 10.1038/nchem.121.
- 884 [96] Bryan R. Goldsmith, Jacques Esterhuizen, Jin-Xun Liu, Christopher J. Bartel, and Christopher Sut-
885 ton. Machine learning for heterogeneous catalyst design and discovery. *AIChE Journal*, 64(7):
886 2311–2323, 2018. doi: 10.1002/aic.16198.
- 887 [97] Takashi Toyao, Zen Maeno, Satoru Takakusagi, Takashi Kamachi, Ichigaku Takigawa, and Ken-
888 ichi Shimizu. Machine learning for catalysis informatics: Recent applications and prospects. *ACS*
889 *Catalysis*, 10(3):2260–2297, 2019. doi: 10.1021/acscatal.9b04186.
- 890 [98] Kevin Tran and Zachary W. Ulissi. Active learning across intermetallics to guide discovery of
891 electrocatalysts for co2 reduction and h2 evolution. *Nature Catalysis*, 1(9):696–703, 2018. doi:
892 10.1038/s41929-018-0142-1.
- 893 [99] Zachary W. Ulissi, Andrew J. Medford, Thomas Bligaard, and Jens K. Nørskov. To address surface
894 reaction network complexity using scaling relations machine learning and dft calculations. *Nature*
895 *Communications*, 8(1), 2017. doi: 10.1038/ncomms14621.
- 896 [100] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. Finding
897 density functionals with machine learning. *Physical Review Letters*, 108(25), 2012. doi: 10.1103/
898 physrevlett.108.253002.
- 899 [101] John C. Snyder, Matthias Rupp, Katja Hansen, Leo Blooston, Klaus-Robert Müller, and Kieron
900 Burke. Orbital-free bond breaking via machine learning. *The Journal of Chemical Physics*, 139(22),
901 2013. doi: 10.1063/1.4834075.
- 902 [102] Wenjian Liu and Mark R. Hoffmann. ici: Iterative ci toward full ci. *Journal of Chemical Theory and*
903 *Computation*, 12(3):1169–1178, 2016. doi: 10.1021/acs.jctc.5b01099.
- 904 [103] Li Li, John C. Snyder, Isabelle M. Pelaschier, Jessica Huang, Uma-Naresh Niranjan, Paul Duncan,
905 Matthias Rupp, Klaus-Robert Müller, and Kieron Burke. Understanding machine-learned density
906 functionals. *International Journal of Quantum Chemistry*, 116(11):819–833, 2015. doi: 10.1002/
907 qua.25040.
- 908 [104] Li Li, Thomas E. Baker, Steven R. White, and Kieron Burke. Pure density functional for strong
909 correlation and the thermodynamic limit from machine learning. *Physical Review B*, 94(24), 2016.
910 doi: 10.1103/physrevb.94.245129.
- 911 [105] Fahhad H. Alharbi and Sabre Kais. Kinetic energy density for orbital-free density functional calcu-
912 lations by axiomatic approach. *International Journal of Quantum Chemistry*, 117(12), 2017. doi:
913 10.1002/qua.25373.
- 914 [106] Ardavan Farahvash, Chee-Kong Lee, Qiming Sun, Liang Shi, and Adam P. Willard. Machine learn-
915 ing frenkel hamiltonian parameters to accelerate simulations of exciton dynamics. *The Journal of*
916 *Chemical Physics*, 153(7), 2020. doi: 10.1063/5.0016009.
- 917 [107] He Zhang, Siyuan Liu, Jiacheng You, Chang Liu, Shuxin Zheng, Ziheng Lu, Tong Wang, Nanning
918 Zheng, and Bin Shao. Overcoming the barrier of orbital-free density functional theory for molecular
919 systems using deep learning. *Nature Computational Science*, 4(3):210–223, 2024. doi: 10.1038/
920 s43588-024-00605-8.
- 921 [108] Yilin Zhao, Dongbo Zhao, Chunying Rong, Shubin Liu, and Paul W. Ayers. Information theory meets
922 quantum chemistry: A review and perspective. *Entropy*, 27(6):644, 2025. doi: 10.3390/e27060644.
- 923 [109] Yilin Zhao, Dongbo Zhao, Chunying Rong, Shubin Liu, and Paul W. Ayers. Extending the
924 information-theoretic approach from the (one) electron density to the pair density. *The Journal*
925 *of Chemical Physics*, 162(24), 2025. doi: 10.1063/5.0270709.
- 926 [110] Shubin Liu. Steric effect: A quantitative description from density functional theory. *The Journal of*
927 *Chemical Physics*, 126(24), 2007. doi: 10.1063/1.2747247.
- 928
929
930
931
932
933
934

- 935 [111] Shubin Liu, Lianghong Liu, Donghai Yu, Chunying Rong, and Tian Lu. Steric charge. *Physical*
936 *Chemistry Chemical Physics*, 20(3):1408–1420, 2018. doi: 10.1039/c7cp07678a.
- 937
938 [112] Bin Wang, Shubin Liu, Ming Lei, and Frank De Proft. Steric effect and intrinsic electrophilicity
939 and nucleophilicity from conceptual density functional theory and information-theoretic approach
940 as quantitative probes of chemical reactions. *Chemistry – A European Journal*, 30(54), 2024. doi:
941 10.1002/chem.202401295.
- 942 [113] Dongbo Zhao, Yilin Zhao, Xin He, Yunzhi Li, Paul W. Ayers, and Shubin Liu. Accurate and efficient
943 prediction of post-hartree–fock polarizabilities of condensed-phase systems. *Journal of Chemical*
944 *Theory and Computation*, 19(18):6461–6470, 2023. doi: 10.1021/acs.jctc.3c00646.
- 945 [114] Jia Fu, Meng Li, Chunying Rong, Dongbo Zhao, and Shubin Liu. Information-theoretic quantities
946 as effective descriptors of electrophilicity and nucleophilicity in density functional theory. *Journal*
947 *of Molecular Modeling*, 30(10), 2024. doi: 10.1007/s00894-024-06116-7.
- 948
949 [115] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf
950 Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein,
951 David A. Evans, Chia-Chun Hung, Michael O’Neill, David Reiman, Kathryn Tunyasuvunakool,
952 Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland,
953 Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Fig-
954 urnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low,
955 Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisun-
956 daram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Židek, Victor
957 Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure
958 prediction of biomolecular interactions with alphafold3. *Nature*, 630(8016):493–500, 2024. doi:
959 10.1038/s41586-024-07487-w.
- 960 [116] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green,
961 Chongli Qin, Augustin Židek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig
962 Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray
963 Kavukcuoglu, and Demis Hassabis. Improved protein structure prediction using potentials from
964 deep learning. *Nature*, 577(7792):706–710, 2020. doi: 10.1038/s41586-019-1923-7.
- 965 [117] Kathryn Tunyasuvunakool, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Židek,
966 Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, Sameer Velankar, Gerard J. Kley-
967 wegt, Alex Bateman, Richard Evans, Alexander Pritzel, Michael Figurnov, Olaf Ronneberger,
968 Russ Bates, Simon A. A. Kohl, Anna Potapenko, Andrew J. Ballard, Bernardino Romera-Paredes,
969 Stanislav Nikolov, Rishub Jain, Ellen Clancy, David Reiman, Stig Petersen, Andrew W. Senior,
970 Koray Kavukcuoglu, Ewan Birney, Pushmeet Kohli, John Jumper, and Demis Hassabis. Highly ac-
971 curate protein structure prediction for the human proteome. *Nature*, 596(7873):590–596, 2021. doi:
972 10.1038/s41586-021-03828-1.
- 973 [118] Yan Yuan, Yilin Zhao, Linling Lu, Junjie Wang, Jingbo Chen, Shubin Liu, Paul W. Ayers, and
974 Dongbo Zhao. Multiproperty deep learning of the correlation energy of electrons and the physico-
975 chemical properties of molecules. *Journal of Chemical Theory and Computation*, 21(12):5997–6006,
976 2025. doi: 10.1021/acs.jctc.5c00414.
- 977
978
979
980
981
982
983
984
985
986
987
988
989