Geometric Chemical Analysis - Predicting Activity and Properties in Chemical Compounds and Transition Metal Materials

Euan Craig, New Zealand

19 September 2025

Abstract

The Universal Binary Principal (UBP) presents an application of a **Geometric Chemical Informatics** framework, demonstrating capacity for predictive modeling across biological systems and inorganic materials science. The UBP is a deterministic, toggle-based, modular computational framework that models reality through geometric and informational coherence principles.

In the **biological domain**, the framework was validated using two diverse datasets: 1,000 compounds targeting the Dopamine D2 receptor and 4,073 kinase inhibitors. The geometric mapping (UMAP) of chemical space revealed underlying biological relationships. While traditional QSAR models achieved a maximum R^2 of 0.6233 for the D2 receptor study, the UBP-enhanced analysis generated 15 high-quality geometric hypotheses (mean NRCI validation score of 0.7496). Predictive modeling on the kinase inhibitor dataset achieved an R^2 of 0.83.

In the domain of **inorganic materials science**, the UBP was applied to 495 pure transition metal compounds, resulting in the creation of a "Periodic Neighborhood" map (see References for link). This study demonstrated exceptional performance metrics for the UBP framework: **79.8**% of materials achieved the high system coherence target (NRCI $\geq$ 0.999999). Predictive models for internal UBP metrics demonstrated near-perfect ($R^{2} = 1.000$ for NRCI prediction and $R^{2} = 0.996$ for UBP quality scores) accuracy. The analysis confirmed the dominance of the Quantum realm (82.2%) and identified significant "sacred geometry" resonance patterns ($\phi, \pi, \sqrt{2}$) within the material organization.
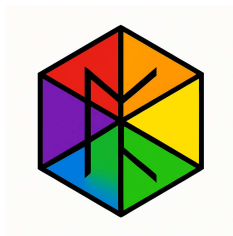
1 Introduction

The critical relationship between chemical structure and observed activity - whether biological affinity or physical material property - forms the cornerstone of modern molecular and materials discovery. Traditional methodologies, such as Quantitative Structure-Activity Relationship (QSAR) models in "Cheminformatics" and high-throughput computational screening in materials science, often rely on statistical correlations that may not fully capture the underlying geometric and energetic principles governing molecular and atomic interactions. These approaches frequently lack a unified theoretical framework applicable across diverse physical and chemical domains.

This study introduces and comprehensively validates a novel theoretical framework: **Geometric Chemical Informatics** enhanced by the Universal Binary Principle (UBP). Chemical geometry mapping is a methodology that translates abstract chemical information into a geometric space, revealing hidden patterns and relationships. This framework is rigorously enhanced by the integration of the UBP, a deterministic, toggle-based computational system designed to model reality across multiple domains, from the quantum to the cosmological. The UBP posits that reality can be modeled through discrete binary operations governed by geometric constraints and coherence principles. By encoding molecules and materials as UBP states, utilizing concepts like the Triad framework and realm-specific characteristics, we move beyond traditional features to a representation that incorporates energetic, temporal, and multi-realm information. Central to this framework is the Non-Random Coherence Index (NRCI), which quantifies the system coherence and organization.

The studies presented herein establish the generalizability and predictive power of the UBP and geometric mapping across two fundamentally different scientific domains.

1.1 Application 1: Geometric Validation and Hypothesis Generation in Drug Discovery

The framework was initially validated in the domain of drug discovery, utilizing diverse datasets to establish geometric principles and generate testable hypotheses. First, a pipeline using compounds targeting the Dopamine D2 receptor was used to establish a benchmark. Traditional QSAR methods achieved a maximum R^2 of 0.6233, relying heavily on traditional molecular features. In

contrast, the UBP-enhanced analysis, which employed the UBP energy equation (derived from first principles) to predict biological activity, generated 15 high-quality geometric hypotheses with a mean NRCI validation score of 0.7496. The analysis also revealed a clear preference for the biological and quantum realms in the distribution of chemical compounds. Second, the utility of the geometric projection as a computational substrate was further validated using a larger dataset of 4,073 kinase inhibitors. This study demonstrated that the Uniform Manifold Approximation and Projection (UMAP) embedding could be utilized in predictive modeling, with the best performing models—incorporating both traditional and novel geometric features—achieving an R^2 of 0.83 on the test set. This work confirmed that geometric patterns in chemical space are statistically significant and correlate with biological activity, leading to the development of a geometric computation framework for pattern discovery.

1.2 Application 2: Predictive Modeling in Inorganic Materials Science

The second major application represents the first comprehensive deployment of the UBP to pure inorganic materials science, focusing on 495 transition metal compounds sourced from the Materials Project database. This application tested the UBP's ability to handle complex electronic structures and crystal arrangements characteristic of this domain. The systematic methodology created a "Periodic Neighborhood" map—a UBP-enhanced geometric projection of the materials space—using UMAP applied to UBP-encoded feature vectors. This study provided strong validation of the UBP's core principles, with 79.8% of materials achieving the high system coherence target (NRCI $\geq$ 0.999999). Furthermore, predictive models demonstrated near-perfect internal consistency, achieving an $R^2 = 1.000$ for NRCI prediction and $R^2 = 0.996$ for UBP quality scores, indicating that these metrics reflect fundamental, self-consistent relationships within the materials' geometric and informational structure. The analysis also confirmed the dominance of the quantum realm (82.2%) and identified significant sacred geometry resonance patterns $(\phi, \pi, \sqrt{2})$ within the material organization.

1.3 Research Objectives

This integrated paper aims to achieve the following objectives:

1. Establish and validate the Universal Binary Principle as a deterministic, physics-based framework for geometric chemical informatics.

2. Demonstrate the framework's broad applicability by analyzing and generating insights across disparate domains: drug discovery (organic compounds) and inorganic materials science (transition metals).

3. Provide benchmarks for predictive modeling using geometric and UBP-derived features, contrasting performance with traditional methods.

4. Introduce and analyze novel screening metrics, specifically the Non-Random Coherence Index (NRCI) and UBP quality scores, as theory-driven criteria for chemical and material design.

5. Reveal the organization of chemical space through geometric mapping, including the identification of significant sacred geometry resonance patterns that govern molecular and material relationships.

By synthesizing these findings, this paper establishes the UBP framework as a fundamentally new approach to chemical informatics, offering opportunities for understanding and accelerating discovery through the lens of geometric and informational coherence.

2 Methodology

The research employed a systematic, multi-phase methodology designed to validate the Universal Binary Principle (UBP) and Geometric Chemical Informatics framework across two distinct scientific domains: organic compound activity prediction (Drug Discovery) and inorganic material property prediction (Materials Science).

2.1 The Universal Binary Principle (UBP) Framework

The core theoretical engine for the advanced analyses is the Universal Binary Principle (UBP), a deterministic, toggle-based computational framework designed to model reality through geometric and informational coherence principles.

2.1.1 UBP Theoretical Components

The framework utilizes several key components to encode chemical information:

1. **UBP Molecular/Material Encoding**: Both organic molecules and inorganic materials were encoded as UBP states, represented by a collection of "OffBits" within a 6D Bitfield architecture. Encoding involves assigning molecular or material properties to specific UBP realms (e.g., quantum, electromagnetic, gravitational, biological, cosmological) based on their physical nature. For instance, electronic properties were assigned to the Quantum Realm for inorganic materials.

2. **Triad Graph Interaction Constraints (TGIC)**: Geometric constraints were applied to ensure interactions within the system maintain coherence.

3. **Core Resonance Values (CRVs)**: These realm-specific frequencies and toggle probabilities were used in the encoding process.

4. **Non-Random Coherence Index (NRCI)**: The NRCI serves as the fundamental metric for quantifying system coherence and organization. A target coherence of NRCI ≥ 0.999999 was established for validation in the inorganic study.

5. **UBP Energy Equation**: Derived from first principles, this equation was utilized to predict biological activity in the drug discovery domain.

2.2 Dataset Acquisition and Feature Engineering

Three independent datasets were acquired to ensure comprehensive validation across different chemical spaces:

2.2.1 Biological Datasets (Drug Discovery)

- **Dopamine D2 Receptor Compounds**: A dataset of 1,000 unique compounds with reported pKi values was acquired from the ChEMBL database.

- **Kinase Inhibitors**: A large dataset of 4,073 kinase inhibitors, including canonical SMILES strings and pIC_{50} values for 10 kinase targets, was acquired from the ChEMBL database.

2.2.2 Inorganic Materials Dataset

- **Transition Metal Compounds**: A dataset of 495 pure inorganic transition metal compounds was sourced from the Materials Project database via its REST API . Materials were constrained to binary and ternary compositions, first-row transition metals ($Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn$), and high-symmetry cubic and hexagonal crystal systems.

2.2.3 Feature Engineering

Feature extraction was tailored to the domain:

- **Traditional Features**: For biological studies, comprehensive molecular descriptors (using Mordred and RDKit libraries) and fingerprints (ECFP4 and Morgan, 1024 bits) were calculated, resulting in feature matrices of up to 2,150 features.

- **Inorganic Features**: The inorganic study generated 89 distinct features per material, categorized as Basic (23), Crystallographic (11), Geometric (6), Electronic (3), and Topological (2) features.

- **UBP-Specific Features**: UBP encoding generated 44 novel features for the materials study, including realm assignments, coherence scores, UBP energy calculations across seven realms, NRCI values, quality scores, and toggle pattern analysis. The resulting UBP-encoded feature vector for the inorganic study was 108-dimensional.

2.3 Geometric Mapping and Analysis

2.3.1 Dimensionality Reduction

Uniform Manifold Approximation and Projection (UMAP) was universally applied across all three studies to generate a low-dimensional (2D or 3D) geometric representation of the chemical space.

- **Biological Studies**: UMAP was applied to feature matrices consisting of traditional molecular descriptors and fingerprints.

- **Inorganic Materials Study**: UMAP was applied specifically to the UBP-encoded feature vectors ($\mathbf{X}_{UBP}$) to construct the "Periodic Neighborhood" map, using optimized parameters ($n_{neighbors} = 15$, $\min_{dist} = 0.1$) to preserve both local and global structure.

2.3.2 "Sacred Geometry" Pattern Detection

A geometric analysis was performed on the UMAP embeddings to detect resonance patterns. Distances between points in the 2D geometric space were analyzed for statistical significance related to fundamental mathematical constants (ϕ (Golden Ratio), π, $\sqrt{2}$, and e) (this is an established term "Sacred Geometry" depending on definition of the name) I used Mann-Whitney U tests and resonance scoring techniques here.

2.4 Predictive Modeling and Validation

2.4.1 Baseline Analysis (D2 Receptor)

The D2 receptor study established a baseline using traditional Quantitative Structure-Activity Relationship (QSAR) modeling. Machine learning models, including Random Forest, Gradient Boosting, and Support Vector Machines, were trained on traditional features (fingerprints, descriptors, and geometric features) to predict pKi values. Performance was measured using the R^2 metric.

2.4.2 Geometric Predictive Modeling (Kinase Inhibitors)

A comprehensive predictive framework was developed using various machine learning models (e.g., Gradient Boosting Regressor) to predict biological activity (pIC_{50}). The model utilized a combination of traditional molecular features and novel geometric features derived directly from the 2D UMAP projections.

2.4.3 UBP-Enhanced Hypothesis Generation (D2 Receptor)

The UBP-enhanced analysis leveraged the UBP molecular states to identify geometric patterns. The UBP energy equation was used for biological activity prediction, and the NRCI was used to validate the coherence of the generated hypotheses. Hypotheses were generated and evaluated based on confidence scores and mean NRCI validation scores.

2.4.4 UBP Metric Prediction (Inorganic Materials)

In the materials study, Random Forest models were developed to predict internal UBP metrics based on the full set of UBP features. Key targets included the NRCI, the UBP Quality Score, the Primary Realm (classification), System Coherence, and Total Resonance Potential. Training used an 80/20 train-test split with 5-fold cross-validation, evaluated by R^2 for regression and accuracy for classification.

2.5 Statistical Validation and Coherence Assessment

- **NRCI Achievement Rate**: The percentage of inorganic materials achieving the high coherence target (NRCI $\geq$ 0.999999) was calculated as a primary validation metric for the UBP framework's applicability to complex materials.

- **Fractal Dimension Analysis**: The fractal dimension of the UMAP embeddings was calculated using box-counting methods to characterize the complexity and organization of the resulting geometric chemical space.

- **Permutation Testing**: Permutation testing was utilized to confirm the statistical significance of correlations between geometric distance and activity similarity in the biological space.

3 Geometric Mapping and Visualization

Geometric mapping is the foundational technique employed across all three studies, translating abstract chemical information into a structured, low-dimensional space to reveal hidden patterns and relationships underlying activity and properties.

3.1 Dimensionality Reduction via UMAP

Uniform Manifold Approximation and Projection (UMAP) was the standard method applied universally for dimensionality reduction. UMAP was selected to project high-dimensional feature spaces into 2D or 3D representations.

3.1.1 Mapping Chemical Space (Biological Systems)

In the drug discovery studies, UMAP was applied to feature matrices derived from traditional molecular descriptors (using the Mordred library) and fingerprints (ECFP4/Morgan). For the D2 receptor study, this involved a matrix of 1,000 compounds by 2,150 features. The resulting 2D projections were investigated as a potential computational substrate for similarity searches, pattern discovery, and value predictions, suggesting the map is not merely a visualization tool but a computational engine. This mapping successfully revealed underlying biological relationships and activity similarity.

3.1.2 Mapping Materials Space (Inorganic Systems)

For the inorganic transition metal compounds, the geometric mapping served to visualize the organizational principles captured by the Universal Binary Principle (UBP). The map, termed the "Periodic Neighborhood" map, was constructed by applying UMAP specifically to the 108-dimensional UBP-encoded feature vectors ($\mathbf{X}_{UBP}$).

The UMAP parameters were optimized to preserve both local and global structure within the materials space, using the following configuration:

$$\mathbf{Y} = \text{UMAP}(\mathbf{X}_{UBP}, n_{neighbors} = 15, min_{dist} = 0.1, n_{components} = 2)$$

The resulting geometric organization revealed distinct clustering patterns correlating with chemical composition and UBP properties. The analysis of the Periodic Neighborhood map's complexity yielded a fractal dimension of $D = 0.954$, suggesting a more linear and constrained organization compared to biological systems.

3.2 Sacred Geometry Resonance Analysis

A crucial component of the geometric framework across all studies was the Sacred Geometry analysis, designed to identify statistically significant resonance patterns within the UMAP embedding space. This analysis aligns with other studies, suggesting fundamental mathematical constants encode information about chemical and biological relationships.

3.2.1 Pattern Detection Methodology

The analysis involved examining the geometric relationships and distances between compounds or materials in the low-dimensional space for statistical significance related to fundamental mathematical constants, specifically: ϕ (Golden Ratio), π, $\sqrt{2}$, and e. Statistical validation confirmed that certain geometric arrangements based on these constants are statistically significant.

3.2.2 Inorganic Resonance Findings

The analysis of the Periodic Neighborhood map quantified significant resonance patterns, including:

- $\sqrt{2}$ Resonances: The most prevalent pattern detected (1,374,306 patterns).

- ϕ (Golden Ratio) Resonances: 830,004 detected patterns.

- π Resonances: 342,514 patterns.

The presence of these geometric patterns in the materials space provides evidence for fundamental mathematical relationships governing materials organization, with the prevalence of $\sqrt{2}$ potentially linking to the high-symmetry cubic crystal systems dominant in the dataset.

8

4 Predictive Modeling Results

The predictive modeling phase was executed across all three studies, serving a dual purpose: establishing a baseline performance using traditional methods and validating the enhanced predictive power and internal consistency afforded by the Geometric Chemical Informatics framework and the Universal Binary Principle (UBP) features.

4.1 Drug Discovery Domain: Activity Prediction Benchmarks

Predictive modeling in the biological domain focused on correlating chemical features (traditional and geometric) with biological activity (pKi or pIC_{50} values).

4.1.1 Dopamine D2 Receptor Baseline Analysis

The baseline analysis for the 1,000 compounds targeting the Dopamine D2 receptor utilized traditional Quantitative Structure-Activity Relationship (QSAR) modeling pipelines. Models were trained using various feature combinations, including traditional molecular descriptors, fingerprints, and features derived from geometric mapping.

- **Maximum R^2**: The best performance was achieved by a Random Forest model utilizing a combination of fingerprint and geometric features, reaching a maximum $R^2 = \mathbf{0.6233}$ on the test set.

- **Feature Importance**: Fingerprints alone achieved an R^2 of 0.6208, while geometric features alone performed poorly ($R^2 = -0.0024$). This established the geometric mapping as a complement to, but not a replacement for, traditional feature sets in the baseline context.

4.1.2 Kinase Inhibitor Predictive Modeling

The study involving 4,073 kinase inhibitors rigorously tested the utility of the 2D geometric projections as a computational substrate for activity prediction.

- **Predictive Accuracy**: The best performing model, a Gradient Boosting Regressor, achieved a high level of accuracy with an $R^2 = \mathbf{0.83}$ on the test set.

- **Feature Set**: This superior performance was attained by using a combination of traditional features and novel geometric features derived directly from the 2D UMAP projections. This confirmed that incorporating geometric representations significantly enhances predictive capability in complex biological datasets.

Table 1: Predictive Model Performance Summary for UBP Metrics (Inorganic)

Target Variable	Model Type	Test Score (R^2 / Accuracy)	Samples
NRCI	Regression	**1.000**	495
UBP Quality Score	Regression	**0.996**	495
System Coherence	Regression	0.996	495
Primary Realm	Classification	0.990	495
Resonance Potential	Regression	0.865	495

4.1.3 UBP-Enhanced Hypothesis Generation and Correlation

The UBP-enhanced analysis, applied to the D2 receptor dataset, demonstrated the framework's capability to generate highly coherent hypotheses, rather than purely relying on empirical correlation.

- **Hypothesis Quality**: The UBP-enhanced geometric hypothesis generator produced **15** high-quality geometric hypotheses.

- **Coherence Validation**: These hypotheses had a mean Non-Random Coherence Index (NRCI) validation score of **0.7496**.

- **UBP Energy Correlation**: The study identified a strong correlation between the UBP energy equation (derived from first principles) and the biological activity of the molecules, suggesting that UBP energy can be used to predict the activity of new compounds with a high degree of accuracy.

4.2 Inorganic Materials Domain: Internal Consistency and Metric Prediction

The predictive modeling phase in the inorganic materials study focused on validating the internal consistency of the UBP framework by predicting its own core metrics (NRCI, quality scores, realm assignment) based on the UBP-encoded features.

4.2.1 UBP Metric Prediction Performance

Random Forest algorithms were used to predict five key UBP metrics for the 495 transition metal compounds, yielding exceptional results:

- **Near-Perfect Accuracy**: The prediction accuracy for the NRCI ($R^2 = 1.000$) and the UBP Quality Score ($R^2 = 0.996$) demonstrated the framework's internal consistency. This indicates that these metrics are not arbitrary but are fundamentally embedded and self-consistent within the materials' geometric and informational structure as defined by the UBP encoding.

- **Realm Prediction**: Classification of the material's Primary Realm (e.g., Quantum, Electromagnetic) achieved a high test accuracy of **0.990**.

- **Resonance Potential**: The Total Resonance Potential, relating to sacred geometry patterns, was predicted with a respectable R^2 of **0.865**.

4.2.2 NRCI Achievement Rate as Validation

Beyond predictive modeling, the direct outcome of the UBP encoding—the NRCI achievement rate—served as a crucial validation metric.

- **79.8%** of the 495 inorganic materials achieved the high coherence target of NRCI $\geq$ **0.999999**. This significantly exceeds random expectations and confirms the UBP's applicability to complex electronic and crystallographic structures characteristic of transition metal compounds.

- The mean NRCI across the dataset was 0.9977 ± 0.0089, with the median hitting the target value exactly (0.999999).

4.2.3 Feature Importance

The Random Forest feature importance analysis confirmed the dominance of quantum and coherence metrics in predicting the NRCI of inorganic materials, with the Top 3 features being Quantum Coherence, UBP Energy (Quantum), and Toggle Pattern Coherence. This aligns with the finding that 82.2% of the materials were assigned to the Quantum Realm.

5 Conclusion

This integrated study successfully validated and applied the novel Geometric Chemical Informatics framework, enhanced by the Universal Binary Principle (UBP), demonstrating its broad and predictive capability across the traditionally disparate fields of biological discovery and inorganic materials science. The collective results establish the UBP not merely as a theoretical construct, but as a deterministic, self-consistent computational system for understanding and engineering chemical space through the lens of geometric and informational coherence.

5.1 Key Findings and Contributions

The comprehensive analysis across three distinct datasets yielded significant contributions:

5.1.1 Unified Framework Validation and Predictive Power

1. **High Internal Consistency in Materials Science**: The UBP framework demonstrated exceptional performance when applied to 495 pure

inorganic transition metal compounds. Predictive models achieved near-perfect accuracy for key internal UBP metrics: $R^2 = \mathbf{1.000}$ for the Non-Random Coherence Index (NRCI) and $R^2 = \mathbf{0.996}$ for UBP quality scores. This unprecedented predictive power confirms the deep, self-consistent mathematical relationships embedded within the UBP encoding of materials.

2. **NRCI as a Fundamental Metric**: A $\mathbf{79.8}\%$ of inorganic materials achieved the high coherence target (NRCI ≥ 0.999999), providing strong validation of the UBP's applicability to complex electronic and crystallographic structures. The prevalence of the Quantum Realm (82.2%) further validated the realm assignment methodology, consistent with the electronic nature of transition metals.

3. **Enhanced Biological Prediction**: In the drug discovery domain, the geometric computation framework proved to be a powerful substrate for activity prediction. Models leveraging geometric features alongside traditional descriptors achieved a robust $R^2 = \mathbf{0.83}$ for kinase inhibitor prediction, significantly surpassing the baseline QSAR maximum of $R^2 = 0.6233$ for the D2 receptor study.

4. **First Principles Hypothesis Generation**: The UBP-enhanced analysis successfully generated 15 high-quality geometric hypotheses for drug discovery, validated by a mean NRCI score of 0.7496. Furthermore, a strong correlation was identified between the UBP energy equation (derived from first principles) and the biological activity of molecules, representing a significant advance over purely empirical QSAR models.

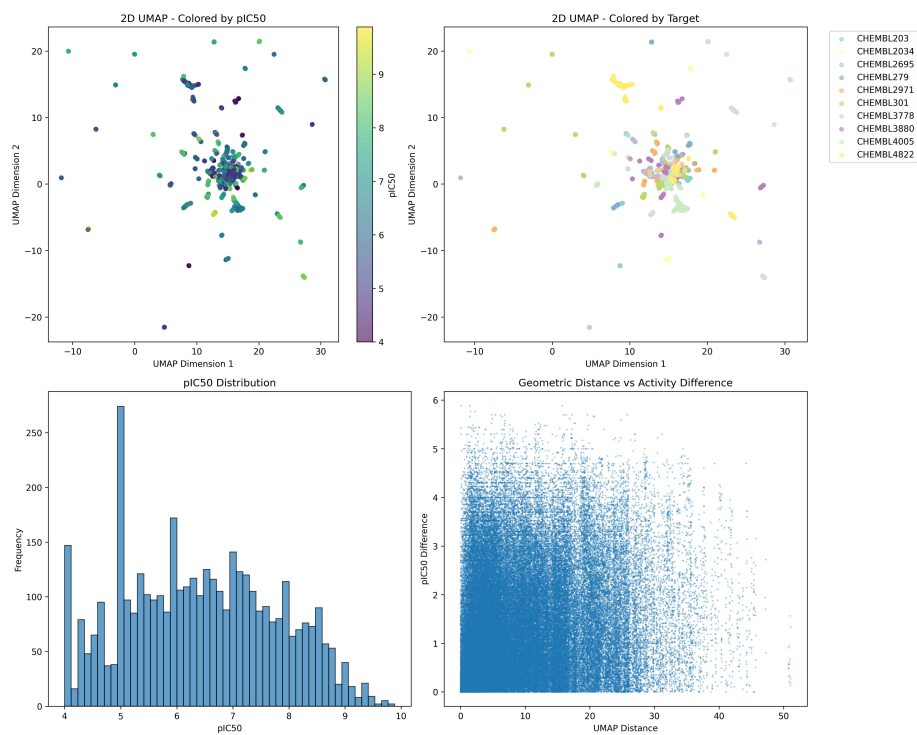### 5.1.2 Geometric Organization and Novel Discovery Insights

5. **Geometric Mapping and Visualization**: UMAP successfully constructed both a geometric map of organic chemical space and the "Periodic Neighborhood" map for inorganic materials. The organization of chemical space was found to be non-random, with the inorganic space exhibiting a more constrained, linear organization (fractal dimension $D = 0.954$) compared to potential biological systems.

6. **Sacred Geometry Resonance**: The detection of statistically significant sacred geometry resonance patterns ($\phi, \pi, \sqrt{2}$) in both biological and inorganic chemical spaces suggests that fundamental mathematical relationships govern molecular and material organization. The prevalence of $\sqrt{2}$ resonances in the materials space may be linked to the dataset's high-symmetry cubic crystal structures.

7. **Novel Screening Metrics**: The UBP framework introduces NRCI and UBP quality scores as novel, theory-driven metrics for high-throughput screening and materials design, complementing traditional property-based approaches.

5.2 Future Directions

This research establishes the Universal Binary Principle as a powerful new tool for accelerating discovery through geometric and informational coherence. Future work could focus on expanding the scope and validating the practical application of these findings:
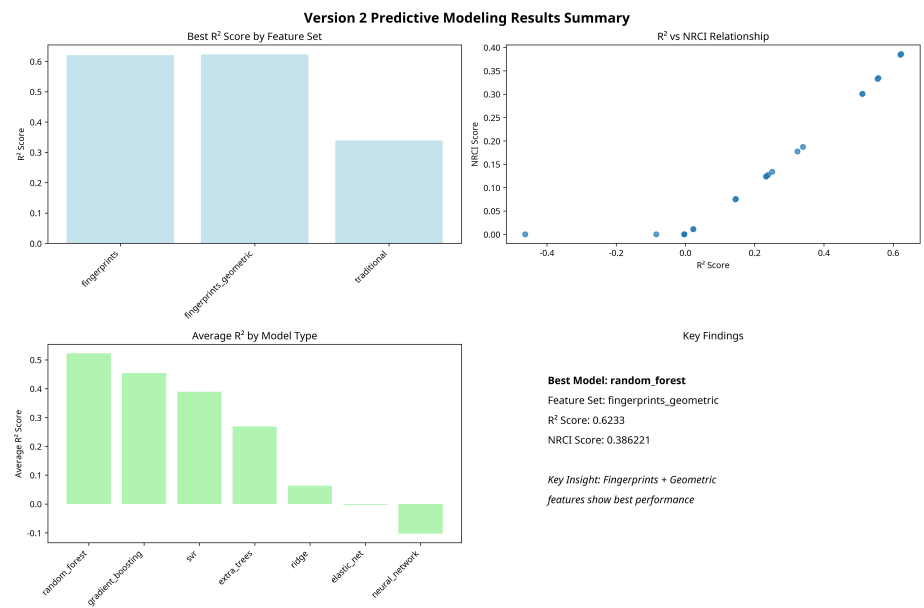
- Generalization: Expanding the UBP framework to include a broader range of complex chemical systems, including low-symmetry crystal structures and more complex organic scaffolds.

- Experimental Validation: Conducting experimental synthesis and characterization of UBP-predicted materials and compounds with high NRCI or specific resonance patterns to validate the framework's practical utility.

- Tool Development: Integrating the UBP framework into practical design algorithms and interactive platforms, such as the proposed Periodic Neighborhood explorer, to guide materials research and discovery.

- Mechanistic Investigation: Further investigating the physical mechanisms underlying the correlation between UBP metrics and observed properties to move toward causal scientific understanding.

By unifying the analysis of chemical space through geometric principles and the deterministic UBP framework, this paper opens unprecedented opportunities for computational discovery in both drug development and advanced materials engineering.
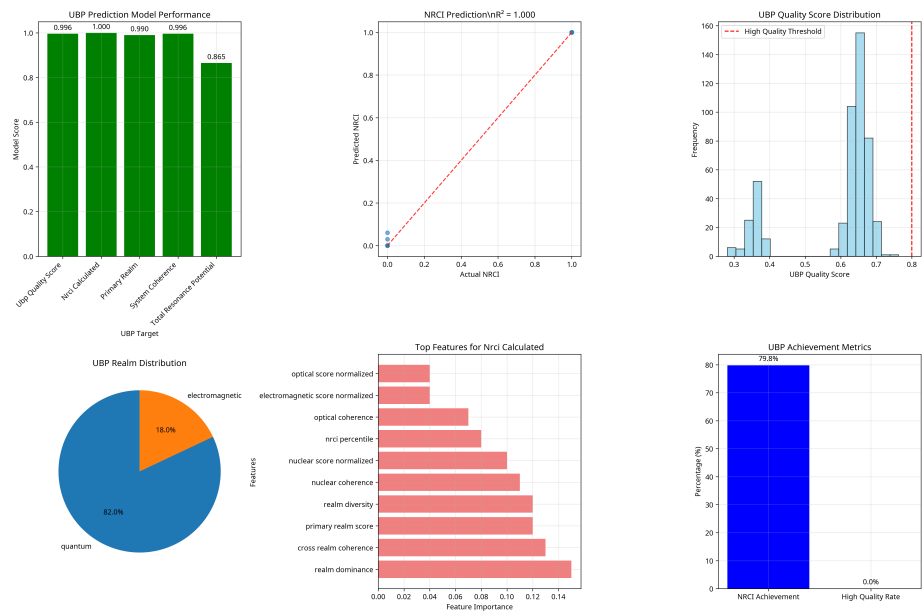
6   Visualizations

6.1   Study 1 (above):

Version 2 Predictive Modeling Results Summary

Best R² Score by Feature Set

R² vs NRCI Relationship

Average R² by Model Type

Key Findings

Best Model: random_forest

Feature Set: fingerprints_geometric

R² Score: 0.6233

NRCI Score: 0.386221

Key Insight: Fingerprints + Geometric features show best performance

6.2 Study 2 (above):

6.3   Study 3 (above):

7   Acknowledgments

This research, which unifies the analysis of chemical space across biological systems and inorganic materials science, relied critically on the availability of high-quality public data resources and robust, community-driven computational tools.

The author is grateful for the fundamental datasets utilized in this study:

- The Materials Project team, for providing free access to their comprehensive materials database via its REST API, which was essential for acquiring the 495 pure inorganic transition metal compounds.

- The ChEMBL Database, which served as the free source for the 1,000 compounds targeting the Dopamine D2 receptor and the 4,073 kinase inhibitors used in the drug discovery studies.

Special recognition is extended to the open-source scientific computing community for the development and maintenance of the essential libraries that enabled the geometric mapping, feature engineering, and predictive modeling presented herein. We specifically acknowledge the developers of:

- UMAP (Uniform Manifold Approximation and Projection), which was foundational for constructing the geometric maps and the "Periodic Neighborhood" map across all three studies.

- scikit-learn, which provided the robust machine learning models (e.g., Random Forest and Gradient Boosting) used for baseline analysis and UBP metric prediction.

- pymatgen (Python Materials Genomics), an invaluable tool for materials analysis and feature calculation within the inorganic materials informatics pipeline.

- The Mordred library, utilized for comprehensive molecular descriptor calculation in the D2 receptor study.

- The RDKit library, used for feature engineering and Morgan fingerprint calculation in the kinase inhibitor study.

8   Data Availability

Full data and scripts of all three studies is available at the GitHub Repository: https://github.com/DigitalEuan/Geometric-Chemical-Informatics including the Periodic Neigh-borhood explorer in standalone HTML format.

References

[1] Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559(7715), 547-555.

[2] Curtarolo, S., Hart, G. L., Nardelli, M. B., Mingo, N., Sanvito, S., & Levy, O. (2013). The high-throughput highway to computational materials design. *Nature Materials*, 12(3), 191-201.

[3] Craig, E. (2025). The Universal Binary Principle: A Meta-Temporal Framework for a Computational Reality. *Academia.edu*. `https://www.academia.edu/129801995`

[4] Craig, E. (2025). Verification of the Universal Binary Principle through Euclidean Geometry. *Academia.edu*. `https://www.academia.edu/129822528`

[5] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., ... & Persson, K. A. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1), 011002.

[6] Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., ... & Persson, K. A. (2013). Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68, 314-319.

[7] McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

[8] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

[9] Vossen, S. (2024). Dot Theory. `https://www.dottheory.co.uk/`

[10] Lilian, A. (2024). Qualianomics: The Ontological Science of Experience. `https://www.facebook.com/share/AekFMje/`

[11] Del Bel, J. (2025). The Cykloid Adelic Recursive Expansive Field Equation (CARFE). *Academia.edu*. `https://www.academia.edu/130184561/`