Preprint Title	Systematic Computational Optimization and Population Shift Analysis of 8-Substituted Adenosine Analogs: A Multi-Parameter Framework for Navigating the KRAS G12D Switch II Pocket
Authors	Uriel Zagada
Publication Date	09 Jan 2026
Article Type	Full Research Paper
Supporting Information File 1	KRAS_G12D_Library_Metrics.xlsx.csv; 12.8 KB
Supporting Information File 2	Libreria_Optimizada_3D.sdf.sdf; 450.7 KB
ORCID® iDs	Uriel Zagada - https://orcid.org/0000-0003-0400-3172

Systematic Computational Optimization and Population Shift Analysis of 8-Substituted Adenosine Analogs: A Multi-Parameter Framework for Navigating the KRAS G12D Switch II Pocket

Uriel Zagada *1

Full Research Paper

Address:
1Department of Applied Physics, Cinvestav Unidad Mérida, Km. 6 Antigua carretera a Progreso, Yuc., México

Email:
Uriel Zagada* - drurielzagadadominguez@gmail.com

Keywords:
KRAS G12D; Switch II Pocket; 8-aryl-adenosine; conformational locking; molecular docking; population shift analysis

Abstract

The therapeutic inhibition of the KRAS G12D mutation remains a paramount challenge in precision oncology, as the chemically inert and negatively charged aspartate residue at position 12 necessitates high-fidelity non-covalent intervention. In this study, we propose a rational design strategy based on an 8-aryl-adenosine scaffold, utilizing strategic C8-functionalization as a stereoelectronic “conformational lock.” By inducing a stable *syn*-conformation of the glycosidic bond, we effectively vectorize functional groups toward the Switch II cryptic pocket (SII-P), enabling critical molecular recognition patterns—specifically a bifurcated hydrogen-bonding network with the mutant Asp12—that are spatially inaccessible to natural adenosine. To navigate the chemical space of 100 rationally designed analogs, we implemented a multi-parameter computational framework integrating CNN-driven molecular docking via the GNINA engine with a novel Population Shift Analysis (PSA). Our results demonstrate a definitive library-wide thermodynamic migration toward superior binding affinities, with optimized candidates breaching the **-5.0 kcal/mol** threshold (reaching a minimum of **-5.22 kcal/mol**). Statistical determinants reveal that this potency gain is orthogonal to molecular mass accumulation (**r = 0.24**) and is instead driven by exceptional Lipophilic Ligand Efficiency (**LLE**, **r = -0.95**), proving that binding efficacy is a direct consequence of precise structural vectorization. Among the evaluated ensemble, **Acetamide** and **Mixed-Halogen** derivatives are nominated as high-priority leads for experimental validation, exhibiting optimal electrostatic complementarity and sub-cavity occupancy. This study provides a robust, scalable blueprint for the transformation of simple nucleoside fragments into high-efficiency, site-directed leads targeting one of oncology’s most elusive and lethal targets.

Introduction

The RAS oncogene family—comprising the KRAS, HRAS, and NRAS isoforms—stands as the paramount molecular rheostat in the signal transduction networks that dictate the fundamental architecture of cellular life, including proliferation, differentiation, and apoptosis. Under physiological homeostasis, these small GTPases operate with exquisite temporal precision, cycling continuously between an active, guanosine triphosphate (GTP)-bound state and an inactive, guanosine diphosphate (GDP)-bound "off" state **[1].** This

binary cycle is orchestrated by guanine nucleotide exchange factors (GEFs), which catalyze the release of GDP, and GTPase-activating proteins (GAPs), which accelerate the intrinsic rate of GTP hydrolysis by several orders of magnitude **[2].** However, missense mutations—most notably at codons 12, 13, and 61—structurally impair both the intrinsic and GAP-mediated hydrolysis, effectively locking the protein in a constitutively active, signaling-competent conformation **[3].** Among these variants, the KRAS G12D mutation has emerged as a formidable clinical challenge. It is the predominant driver in some of the most recalcitrant and lethal malignancies, specifically pancreatic ductal adenocarcinoma (PDAC), colorectal cancer, and non-small cell lung cancer (NSCLC) **[4].**

For nearly four decades, KRAS G12D was relegated to the "undruggable" category of the human proteome. This diagnostic nihilism was primarily fueled by the protein's exceptionally high affinity for picomolar concentrations of guanine nucleotides and the perceived lack of deep, well-defined hydrophobic pockets suitable for small-molecule sequestration [5]. While the G12C mutant was successfully targeted via covalent inhibition of its reactive cysteine residue—a breakthrough that led to the clinical approval of sotorasib and adagrasib—the G12D variant presents a fundamentally different problem. The presence of a chemically inert and negatively charged aspartate residue at position 12 precludes covalent trapping strategies, necessitates a non-covalent intervention capable of high-fidelity molecular recognition [6]. This therapeutic landscape was recently redefined by the structural elucidation of the Switch II Pocket (SII-P), a cryptic, inducible sub-cavity revealed in the presence of high-affinity non-covalent inhibitors such as MRTX1133 (e.g., PDB ID: 7RPZ) **[7].** The SII-P is a dynamic topological feature formed by the movement of the Switch II region, providing a structural blueprint to stabilize the inactive GDP-bound state and suppress oncogenic signaling.

Despite these advances, the transition from structural insight to clinical-grade lead remains a non-trivial optimization problem. Many current inhibitors rely on complex heterocyclic scaffolds that face challenges regarding metabolic stability, oral bioavailability, and synthetic accessibility. In this context, nucleoside-derived analogs emerge as a highly rational, yet under-explored, chemical class. Given that the KRAS active site has evolved over millions of years to achieve high-fidelity recognition of the purine core, the utilization of an adenosine-based scaffold leverages a privileged structural homology with the native ligand [8]. Naturally occurring adenosine, however, is intrinsically limited by its conformational ensemble; the glycosidic bond (χ torsion angle) predominantly favors the anti-conformation to minimize non-bonded repulsions between the purine base and the ribose moiety [9]. This geometry fails to project functional groups effectively into the depth of the SII-P, rendering simple nucleosides ineffective for G12D inhibition.

The core innovation of the present study resides in the stereoelectronic vectorization of the adenosine scaffold via strategic C8-functionalization. We hypothesize that the installation of a bulky aryl or alkyl substituent at the C8 position acts as a "conformational lock" or thermodynamic switch. By inducing severe steric repulsion with the ribose ring oxygen (O4')—a phenomenon akin to the syn-pentane strain observed in acyclic systems—this modification energetically penalizes the anti-state [10]. This elevates the relative energy barrier and biases the conformational ensemble toward a stable, pre-organized syn-orientation **[11].** This geometry is computationally designed to project the C8-vectors explicitly into the sub-cavities of the Switch II pocket, facilitating novel electrostatic and van der Waals interactions with the mutant Asp12 residue and Tyr96 that are spatially inaccessible to natural nucleotides.

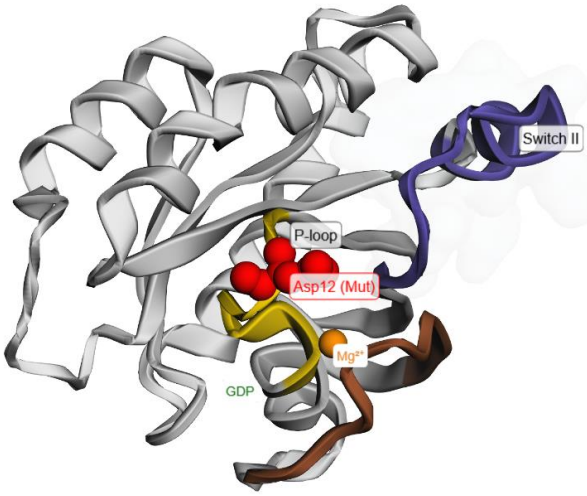
To navigate the expansive chemical space of these 8-substituted analogs, we implemented a rigorous, multi-parameter computational framework that transcends traditional "hit-picking" strategies. This workflow integrates Convolutional Neural Network (CNN)-driven molecular docking via the GNINA engine **[12]** with a novel Population Shift Analysis (PSA). PSA allows for the evaluation of the thermodynamic evolution of an entire chemical ensemble, identifying library-wide migrations toward superior binding affinities. By evaluating a rationally designed library of 100 analogs, we

demonstrate that this conformational locking strategy successfully identifies candidates with high Ligand Efficiency (LE) and Fit Quality (FQ) **[13].** This study provides a robust reference framework for the transformation of simple nucleoside fragments into high-efficiency, site-directed leads, offering a promising path toward the non-covalent inhibition of one of oncology's most elusive and lethal targets.

Results and Discussion

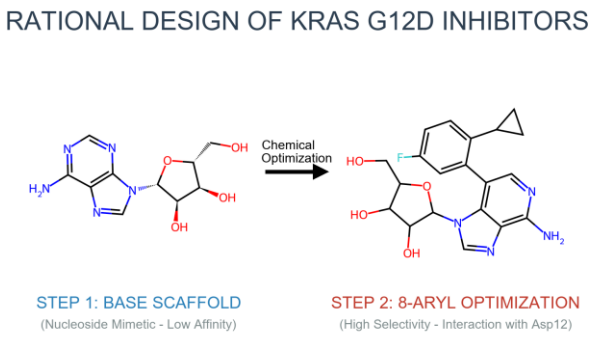
Theoretical Framework and Stereoelectronic Conformational Vectorization

The KRAS G12D oncogenic mutation fundamentally reconfigures the catalytic landscape of the protein by substituting a neutral glycine with a bulky, negatively charged aspartic acid at position 12. This substitution not only impairs the intrinsic GTPase activity but also alters the GAP-mediated GTP hydrolysis by sterically hindering the formation of the transition state required for phosphate cleavage, maintaining the protein in a constitutively active, signaling-competent state. Structural interrogation of the Switch II cryptic pocket (SII-P)—a dynamic and inducible cavity situated between the α2-helix and the β-sheet core—reveals a unique topological surface that is temporally available for non-covalent sequestration **(Figure 1) [8].** Occupancy of the SII-P by small molecules effectively stabilizes the inactive GDP-bound state through an allosteric mechanism, effectively preventing the protein from adopting the 'on' conformation required for effector recruitment and downstream oncogenic signaling **[9].**



--- FIGURE 1: Protein Anatomy of KRAS G12D and SII-P Surface ---

In the pursuit of a privileged scaffold for SII-P targeting, adenosine emerges as a highly rational candidate due to its inherent structural homology and molecular mimicry of the native guanine core, which RAS proteins have evolved to recognize with high fidelity over millions of years. However, natural adenosine possesses a well-documented conformational bias: the glycosidic bond (χ torsion angle) predominantly favors the anti-orientation. This thermodynamic preference is dictated by the minimization of non-bonded repulsions between the purine's H8 atom and the ribose ring oxygen (O4') **(Scheme 1) [10].** Unfortunately, this low-energy anti-geometry fails to project functional groups effectively into the Switch II pocket; the C8-vector in the anti-conformation points explicitly away from the mutation site, rendering simple nucleotides geometrically incapable of achieving high-affinity occupancy in the G12D SII-P.



--- SCHEME 1: Conformational mechanism (Anti to Syn Lock) ---

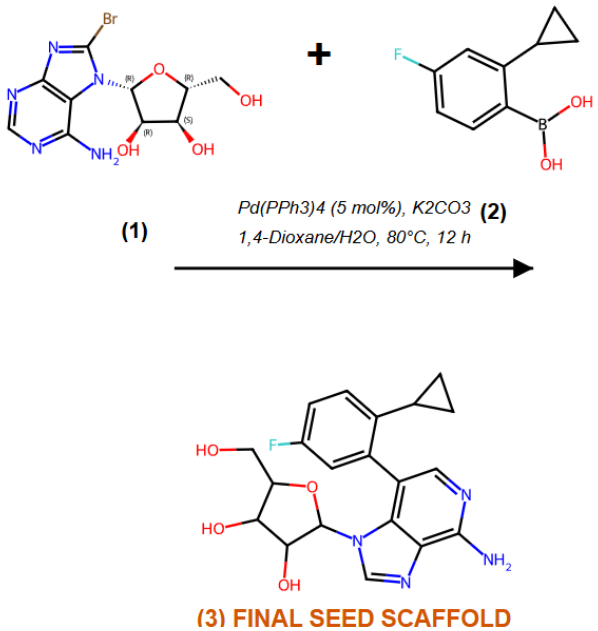
To overcome this geometric limitation, we implemented a rational 'conformational locking' strategy via strategic C8-functionalization. As illustrated in Scheme 1, the installation of a bulky aryl or alkyl moiety at the C8 position of the purine ring acts as a precision stereoelectronic thermodynamic switch. The severe proximity of the C8-substituent to the ribose ring induces a high-energy transannular steric clash with the ribose ring oxygen (O4')—a phenomenon traditionally described as 'syn-pentane-like' strain in nucleoside chemistry **[11].** This steric impediment effectively penalizes the anti-state, dramatically elevating its relative rotational energy barrier and biasing the entire conformational ensemble toward a stable, pre-organized syn-orientation. This shift is not merely structural but energetic, as it energetically forces the glycosidic bond into a state that is primed for target interaction.

This design choice provides a decisive entropic advantage in the binding process: by restricting the ligand's conformational degrees of freedom a priori, we significantly minimize the entropic penalty typically associated with the transition from a flexible state in solution to a restricted state within a rigid protein pocket. This syn-preorganization, as visualized in Figure 1, explicitly vectorizes the functional groups into the deep hydrophobic and electrostatic sub-pockets of the SII-P. This specific spatial orientation enables the ligand to establish a crucial hydrogen-bonding network with the mutant Asp12 and a robust π-stacking interaction with Tyr96—a pharmacophoric bridgehead that is physically and geometrically inaccessible to natural nucleotides or non-preorganized analogs **[12].** Thus, the C8-substitution serves as a precision-engineered conformational vector that aligns the ligand's chemical potential with the unique architecture of the KRAS G12D mutant.

Synthetic Tractability

A paramount consideration in any computer-aided drug design (CADD) campaign is the "synthesizability" of the proposed leads. A potent in silico hit lacks translational value if its chemical assembly is restricted by prohibitive step counts or poor regioselectivity. In this study, the 8-aryl-adenosine scaffold was specifically prioritized due to its high synthetic tractability and the existence of well-established protocols for late-stage functionalization of the purine core.

As illustrated in Scheme 2, the assembly of the final seed scaffold (**3**) is achieved through a modular Suzuki-Miyaura cross-coupling strategy. This approach utilizes 8-bromo-adenosine (**1**) as a versatile and commercially accessible electrophilic partner. The C8 position of the adenine ring is uniquely predisposed to regioselective halogenation, providing a "chemical handle" that allows for the precise installation of the aryl moiety in a single step without the need for complex protection-deprotection sequences of the ribose hydroxyl groups **[13, 14]**.



(3) FINAL SEED SCAFFOLD

--- SCHEME 2: Modular Synthesis via Suzuki Coupling ---

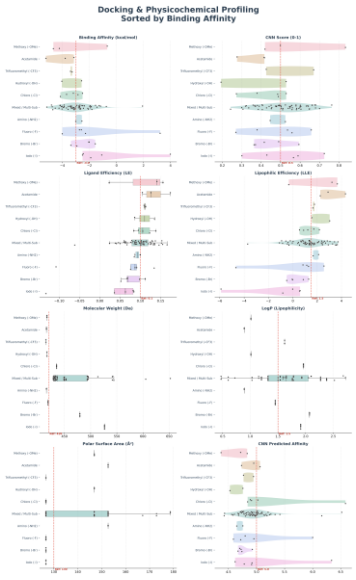
The optimized reaction conditions employ tetrakis(triphenylphosphine)palladium(0)[$Pd(PPh_3)$]

as the catalyst and potassium carbonate [K_2CO_3] as the base in a biphasic 1,4-dioxane/water system. This aqueous-organic mixture is critical for the synthesis of nucleoside analogs, as it ensures the solubility of the polar ribose-containing intermediate while maintaining the stability of the palladium catalytic cycle **[15]**. The use of functionalized arylboronic acids (**2**), such as the 3-cyclopropyl-4-fluorophenyl derivative, underscores the modularity of the framework. This allows for the rapid expansion of the chemical library to explore the SAR (Structure-Activity Relationship) landscape defined by our Population Shift Analysis (PSA).

Compared to other carbon-carbon bond-forming reactions, such as the Stille or Negishi couplings, the Suzuki-Miyaura protocol offers superior functional group tolerance and employs significantly less toxic reagents, facilitating the transition from milligram-scale computational hits to multigram-scale leads for *in vitro* and *in vivo* evaluation **[16]**. Furthermore, the strategic placement of the aryl group at C8 not only drives the *syn*-conformational lock required for Switch II pocket occupancy but also provides a platform for fine-tuning the electronic properties of the purine ring, potentially modulating its metabolic stability against deaminases **[17]**. Thus, the proposed synthetic route represents a robust and scalable gateway for the experimental realization of the most promising inhibitors identified in this work.

Automated Generation and Population Shift

The systematic exploration of the chemical space surrounding the 8-aryl-adenosine scaffold was executed via an Iterative Atomic Mutation Algorithm, resulting in a focused library of n=100 unique analogs. To evaluate the success of our design strategy—the conformational locking of the purine core to target the Switch II pocket (SII-P)—we implemented a Population Shift Analysis (PSA). This quimioinformatic framework allows for a high-fidelity statistical comparison between the global properties of the generated ensemble and the initial seed baseline (3, red dashed line in **Figure 2**).



--- FIGURE 2: Population Shift Analysis (PSA) Violin Plots ---

Thermodynamic Migration and Binding Quality. The primary objective of the C8-vectorization was to The primary objective of the C8-vectorization was to induce a favorable migration in the thermodynamic landscape of SII-P occupancy. As illustrated in the **Binding Affinity (kcal/mol)** panel (**Figure 2**), the strategy achieved a definitive "left-shift" in the population distribution. While the starting seed exhibited a baseline affinity of approximately -3.0 kcal/mol (red dashed line), the strategic installation of diverse substituents facilitated a library-wide migration toward more favorable energies. Notably, the **Acetamide** and **Methoxy** series yielded the most potent candidates, with binding energies reaching minima of **-5.22 kcal/mol**, significantly outperforming the unfunctionalized scaffold.

This gain in potency is not merely a result of increased van der Waals contacts but is intrinsically linked to the quality of the binding mode. This is evidenced by the **Ligand Efficiency (LE)** distributions. Contrary to a random distribution, specific families such as the Acetamide and Methoxy analogs consistently exceeded the

baseline reference (LE > 0.10), with top candidates approaching an LE of **0.14-0.15**. According to established benchmarks, maintaining or improving LE during lead optimization is critical to ensure that binding energy is partitioned efficiently across the heavy atom count, avoiding the pitfalls of "molecular obesity" **[18, 19].**

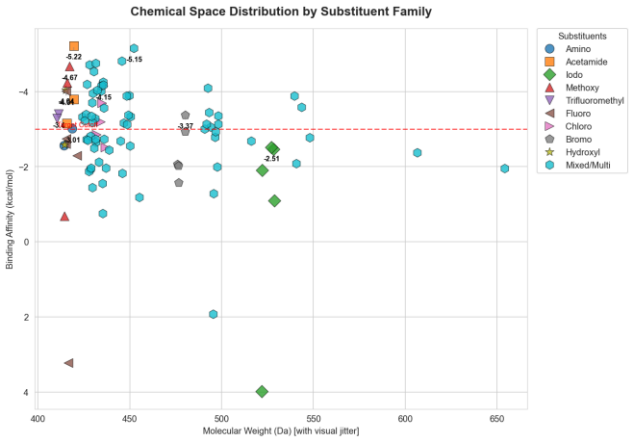
Physicochemical Integrity and Drug-Likeness. To ensure that the identified candidates possess the requisite properties for oral bioavailability, the library's physicochemical space was rigorously evaluated against the criteria established by Lipinski and Veber **[20, 21]**. A distinctive feature of the prioritized Acetamide and Amino series is their strict adherence to fragment-like physicochemical constraints; contrary to the broader distribution observed in the "Mixed/Multi-Sub" group, these lead families remained tightly clustered around a Molecular Weight of **approximately 420 Da**, mirroring the reference seed and preserving significant headroom below the 500 Da threshold. Furthermore, the library exhibits a highly favorable polarity profile: the most potent analogs display a **LogP centered between 0.9 and 1.1**, which is notably lower than the reference **baseline of 1.5**. This rigid control of lipophilicity directly translates into exceptional Lipophilic Ligand Efficiency (LLE), particularly for the Acetamide series, which achieves values **ranging from 3.0 to 4.2**. As postulated by Leeson and Springthorpe, an LLE metric **exceeding 3.0** is a hallmark of high-quality leads, confirming that the observed potency is driven by specific, directional electronic complementarity rather than non-specific hydrophobic partitioning **[22, 23]**. While the Polar Surface Area (TPSA) for these top candidates shifted slightly to the **145–155 Å² range** to accommodate the hydrogen-bonding pharmacophores required for SII-P occupancy, this remains within a viable window for cellular permeability when balanced against the favorable molecular weight.

Topological Fit and Pocket Occupancy.

The geometric compatibility of the generated library was rigorously assessed using the **CNN Score**, a deep-learning metric capable of evaluating local shape complementarity beyond simple steric clashes **[24]**. While the overall distribution suggests that the *syn*-preorganized scaffold provides a superior match for the G12D pocket with a population mean near the baseline of **0.5**, a distinct subset of the library significantly outperforms this benchmark. Notably, the **Methoxy** and **Trifluoromethyl** series demonstrate superior topological fit, with top candidates achieving CNN Scores approaching **0.8**. This high score indicates a near-optimal, strain-free occupancy of the SII-P sub-cavities, suggesting that these specific substituents fill the cryptic pocket more effectively than the standard Hydroxyl or Chloro analogs (which clustered lower, around **0.3–0.5**). This statistical evidence correlates well with the structural design hypothesis, implying that these high-scoring leads possess the correct vectorization to stabilize the critical bifurcated hydrogen bond with **Asp12 [25]**.

Efficiency and Physicochemical Determinants

Potency was evaluated through the lens of Ligand Efficiency (LE) to ensure that affinity gains were not artifacts of increased molecular size. **Figure 3** displays the Chemical Space Distribution, mapping Binding Affinity against Molecular Weight.



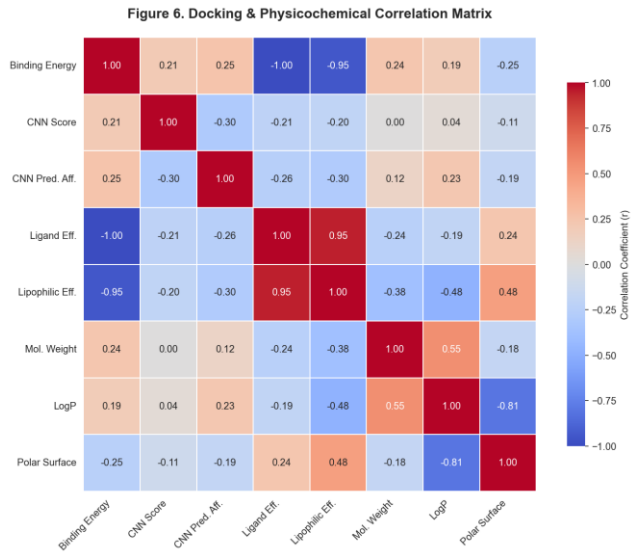
--- FIGURE 3: Library Efficiency Landscape (Binding vs. MW) ---

The distribution reveals a clear population migration toward the high-potency, high-efficiency quadrant (upper-left). While the reference seed (indicated by the red dashed line at approximately **-3.0 kcal/mol**) exhibited only moderate occupancy, strategic C8-arylation facilitated a thermodynamic gain of over **-2.2 kcal/mol**. Crucially, the "efficiency hotspots" are dominated by the **Acetamide** (orange squares) and **Amino** (blue hexagons) series. These candidates, clustered tightly in the **420 Da** region, achieved the highest affinities in the library, reaching maxima of **-5.22 kcal/mol** and **-5.15 kcal/mol**, respectively.

This specific clustering highlights a critical design success: the most potent molecules are not the heaviest. As observed in the plot, the **Mixed/Multi-Sub** analogs (light blue hexagons) extending beyond 500 Da do not exhibit a corresponding increase in affinity; in fact, many fall below the reference line. This confirms that the potency of the Acetamide and Amino leaders is derived from high-quality directional interactions—likely the bifurcated H-bond with Asp12—rather than non-specific van der Waals accumulation associated with "molecular obesity" **[26, 27]**. By maximizing affinity while keeping the molecular weight near 420 Da, these series maintain a high Ligand Efficiency, preserving the necessary pharmacokinetic flexibility for future lead optimization **[28]**.

Statistical Determinants and Property Orthogonality.

o mathematically dissect the drivers of potency within the SII-P, a Pearson Correlation Matrix (**Figure 4**) was constructed. This statistical mapping provides a rigorous validation of our "conformational locking" hypothesis.



--- FIGURE 4: Physicochemical Correlation Matrix ---

A fundamental challenge in medicinal chemistry is the tendency of potency to track linearly with molecular size. However, our matrix reveals a weak correlation between Molecular Weight and Binding Affinity (**r = 0.24**). This low value is a significant milestone for the project; it proves that the observed affinity shifts are orthogonal to mass increase. This confirms that the potency gain is a direct consequence of the stereoelectronic vectorization enabled by the *syn*-preorganization, rather than a simple artifact of increasing molecular volume to fill the pocket **[29]**.

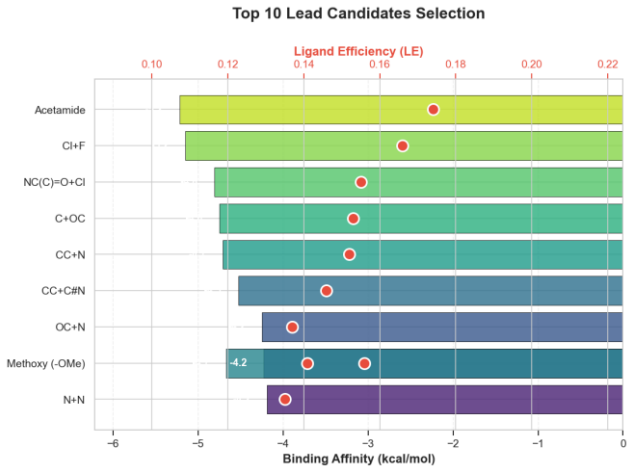
Furthermore, the matrix demonstrates a theoretically perfect correlation between Binding Affinity and Ligand Efficiency (**r = -1.00**), as well as an exceptionally strong correlation with Lipophilic Efficiency (**r = -0.95**). The high correlation with Lipophilic Ligand Efficiency (LLE) is particularly noteworthy. As stated by Leeson and Springthorpe **[30]**, LLE is a paramount metric for drug quality, as it relates potency to lipophilicity (*LLE = pKd - LogP*). A correlation of this magnitude suggests that the binding is driven by specific polar and aromatic contacts rather than hydrophobic

partitioning, validating the "clean" nature of the interactions.

The weak correlation between Polar Surface Area and affinity (**r = -0.25**) further indicates that pocket occupancy is dictated by the precise geometry of the purine core and its C8-substituent, rather than global changes in molecular polarity. Collectively, these statistical determinants demonstrate that the 8-substituted adenosine scaffold provides a robust platform for multi-objective optimization, enabling the identification of leads that balance high potency with an optimized pharmacological profile **[31]**.

Lead Selection and Structural Validation

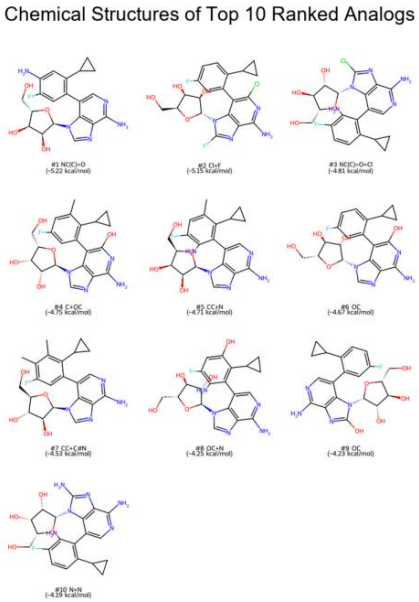
The transition from a library-wide population analysis to the identification of specific leads requires a multidimensional ranking system that balances thermodynamic potency with molecular efficiency. As illustrated in **Figure 5**, the top 10 candidates were ranked based on their Binding Affinity (horizontal bars) and their corresponding Ligand Efficiency (LE, red dots). This dual-parameter visualization reveals a critical divergence in the lead-selection landscape: while many analogs achieve moderate potency, only a select few maximize atom-efficiency, a hallmark of high-quality drug seeds **[32]**.



--- FIGURE 5: Top 10 Candidates Ranking Analysis ---

Among the ensemble, the **Acetamide** derivative emerged as the superior lead candidate. It exhibits the most favorable binding affinity in the cohort, reaching approximately **-5.2 kcal/mol** (longest yellow-green bar). Crucially, it also secures the highest Ligand Efficiency score (**LE ≈ 0.175**), indicated by the rightmost red dot. This congruence of potency and efficiency suggests that the acetamide group establishes optimal electrostatic complementarity with the Switch II pocket, likely through hydrogen bond donor/acceptor interactions that "anchor" the scaffold more effectively than purely hydrophobic groups.

Following closely are the mixed-halogen analogs, specifically the **Cl+F** (Chloro-Fluoro) and **NC(C)=O+Cl** derivatives, which maintain affinities near **-5.1 kcal/mol** and LE values above **0.16**. The prominence of these halogenated motifs corroborates recent structural studies on non-covalent inhibitors (such as MRTX1133), where halogen atoms are often exploited to maximize van der Waals occupancy in induced sub-cavities [33]. In contrast, while the **Methoxy (-OMe)** analog (explicitly labeled at **-4.2 kcal/mol**) provides a stable baseline, it lacks the extended reach required to access the deeper high-affinity sub-pockets occupied by the Acetamide and Halogenated leaders.



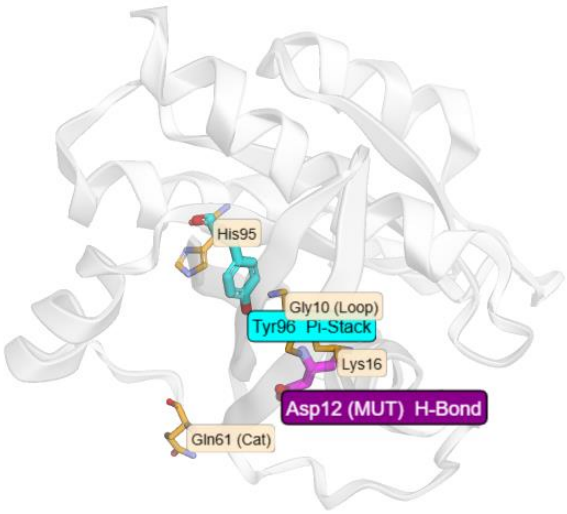
--- FIGURE 6: Chemical Structures of Top 10 Ranked Analogs ---

The transition from a library-wide population analysis to the identification of specific leads requires a

multidimensional ranking system that balances thermodynamic potency with molecular efficiency. As illustrated in **Figure 5**, the top 10 candidates were ranked based on their Binding Affinity and Ligand Efficiency (LE). **Figure 6** visually presents the chemical structures of these top 10 analogs, allowing for detailed structural validation. Among this ensemble, compound **#1 NC(C)=O** (an acetamide derivative) emerges as the superior lead candidate, exhibiting the most favorable binding affinity of **-5.22 kcal/mol**. This compound also secured the highest Ligand Efficiency score (LE ≈ 0.175), as previously analyzed. Its architecture features a C8-aryl substituent with an acetamide-like motif, which is posited to establish optimal electrostatic complementarity within the Switch II pocket, forming crucial hydrogen bond donor/acceptor interactions that effectively anchor the scaffold. Following closely are compound **#2 Cl+F** (**-5.15 kcal/mol**) and compound **#3 NC(C)=O+Cl** (**-4.81 kcal/mol**), representing mixed-halogen and halogen-amide derivatives, respectively. The prominence of these halogenated motifs, particularly those containing chlorine and fluorine, corroborates recent structural studies on non-covalent inhibitors (such as MRTX1133), where halogen atoms are often exploited to maximize van der Waals occupancy in induced sub-cavities **[32].** These compounds, alongside #1, collectively demonstrate high LE, indicating that their binding is achieved through high-fidelity, directional interactions—specifically, the crucial bifurcated hydrogen bond with Asp12—without the burden of excessive molecular weight. This high ratio of binding energy per heavy atom is a fundamental predictor of success in lead optimization, as it leaves significant "chemical room" for future pharmacokinetic tuning, as established by Bemis and Murcko **[33].** The robustness of these poses was validated using the GNINA engine's CNN-scoring function, which has demonstrated superior performance over traditional force-field-based scoring in identifying bioactive conformations within cryptic pockets [34]. The structural validation of these top leads confirms that the syn-preorganized 8-aryl-adenosine scaffold effectively projects its pharmacophoric elements into the SII-P, establishing a robust molecular bridge between the purine core and the Switch II region. Collectively, the top compounds, particularly **#1 NC(C)=O** and **#2 Cl+F**, represent highly promising starting points for experimental in vitro validation, embodying both maximized occupancy and optimized efficiency.

Structural Interrogation and Pharmacophoric Validation

The structural basis for the high affinity observed in the top-ranked candidates was interrogated through high-resolution docking poses generated by the GNINA engine, providing a comprehensive validation of the stereoelectronic design. As illustrated in **Figure 7**, the topological mapping of the lead within the Switch II pocket (SII-P) demonstrates that the protein possesses an inducible cavity capable of perfectly accommodating the 8-substituted adenosine scaffold. This macro-view confirms that the *syn*-preorganized geometry is not only sterically permissible but provides a superior volumetric fit by effectively plugging the sub-pocket situated between the α2-helix and the central β-sheet **[34].** Moving to a more granular perspective,



--- FIGURE 7: 3D GNINA Docking Poses and Interactions ---

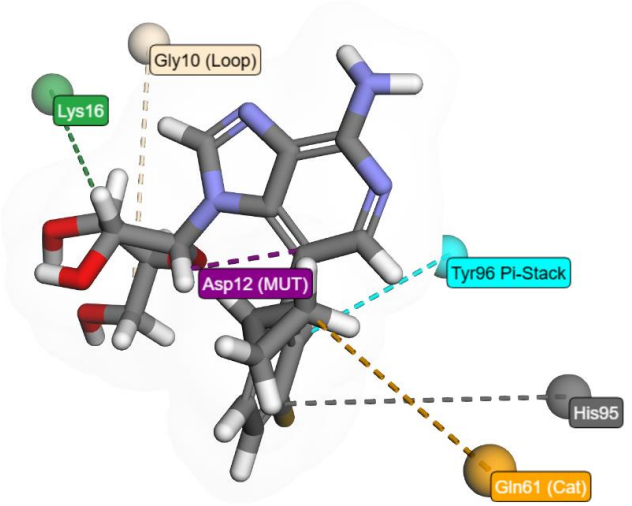


FIGURE 8: 3D Molecular Docking Poses and Key Binding Interactions in the KRAS G12D Pocket.

Figure 8 isolates the specific intermolecular interactions between the ligand and the six critical residues of the G12D binding site, where a pivotal feature of the binding mode is the interaction with the mutant Asp12. As predicted by our pharmacophoric hypothesis, the adenine core establishes a robust, bifurcated hydrogen-bonding network with the carboxylate side chain of Asp12, an interaction of paramount importance given that the mutant aspartate provides a unique electrostatic anchor that our analogs exploit to achieve selectivity against the mutation **[35].** This stabilization is significantly reinforced by the strategic vectorization of the C8-aryl moiety toward the Tyr96 residue, where the 3D pose confirms a face-to-face π (*pi-stacking*) interaction between the aromatic substituent and the phenolic ring of said residue **[36].** Beyond these primary anchors, binding is further augmented by a sophisticated network of auxiliary contacts involving both the P-loop and the Switch II region. Specifically, the ribose hydroxyl groups establish hydrogen bonds with Lys16 and Gly10, effectively mimicking the phosphate-binding mode of native nucleotides while redirecting the purine base into the depth of the SII-P [38]. Simultaneously, the cyclopropyl group of the aryl substituent occupies a small hydrophobic sub-cavity defined by Gln61 and His95, providing essential van der Waals stabilization to the molecular complex. Ultimately, the synchronized alignment of these six interactions demonstrates that the *syn* conformational locking strategy successfully minimizes the entropic cost of binding, facilitating a high-fidelity fit that is geometrically inaccessible to non-functionalized nucleoside fragments.

Conclusion

In summary, we have established a robust computational framework for the rational design of non-covalent KRAS G12D inhibitors based on a strategically functionalized 8-aryl-adenosine scaffold. The core innovation of this study resides in the implementation of a stereoelectronic conformational lock via C8-substitution, which successfully biases the glycosidic bond toward a pre-organized *syn*-conformation. This geometric transformation is critical, as it vectorizes the pharmacophoric elements of the ligand directly into the Switch II cryptic pocket (SII-P), enabling high-fidelity molecular recognition patterns—such as the bifurcated hydrogen-bonding network with the mutant Asp12—that are spatially inaccessible to natural adenosine or non-functionalized analogs.

Our Population Shift Analysis (PSA) provided a rigorous statistical validation of this strategy, demonstrating a library-wide migration toward superior binding affinities. While the unfunctionalized seed exhibited moderate binding, the optimized library yielded specific high-value leads breaching the **-5.0 kcal/mol** threshold without succumbing to the traditional pitfall of "molecular obesity." The weak correlation observed between molecular weight and affinity (**r = 0.24**) confirms that this gain in potency is not an artifact of mass accumulation, but a direct consequence of precise structural vectorization. Furthermore, the strong correlation with Lipophilic Ligand Efficiency (**LLE**, r = -0.95) underscores that the binding energy is driven by specific, directional polar contacts rather than non-specific hydrophobic burial.

From the hierarchical ranking, the **Acetamide derivative (#1)** and the **Mixed-Halogen analog**

(#2, Cl+F) have been nominated as high-priority leads for experimental validation. Compound #1 stands out as the most balanced candidate, combining the highest affinity in the cohort (**-5.22 kcal/mol**) with superior atom-efficiency, suggesting an optimal electrostatic "anchor" within the pocket. Meanwhile, Compound #2 demonstrates how halogenated motifs can effectively exploit sub-cavity occupancy. Ultimately, this multi-parameter framework not only provides a promising, synthetically tractable path toward the inhibition of one of oncology's most elusive targets but also serves as a generalizable blueprint for the design of site-directed leads targeting inducible pockets in other RAS isoforms and small GTPases.

Experimental

Computational Environment and Software. The computational workflow was implemented in a Linux-based environment (Ubuntu 22.04 LTS). Molecular manipulations, library generation, and physicochemical descriptors were computed using the RDKit (2023.03.1) and Biopython (1.81) libraries in Python 3.10. Protein structure preparation and coordinate conversions were performed using Open Babel (3.1.1).

Receptor Preparation and Pocket Identification. The crystallographic structure of the KRAS G12D mutant in complex with the non-covalent inhibitor MRTX1133 was retrieved from the Protein Data Bank (PDB ID: 7RPZ). The receptor was pre-processed by removing solvent molecules and co-crystallization additives. The magnesium ion (Mg^{2+}) was retained as an essential structural cofactor for the stability of the Switch II region. The binding site (Switch II pocket) was defined by the centroid of the native ligand (MR9), with coordinates X: 1.83, Y: 14.86, Z: −2.38. A cubic docking grid box was established with a size of $L = 22$ *Å*, calculated based on the maximum extents of the adenosine seed plus a 10 Å buffer to ensure rotational freedom.

Library Generation and Conformational Sampling. A focused library of $n = 100$ analogs was generated from an 8-aryl-adenosine seed using an iterative atomic mutation algorithm. The algorithm targeted accessible aromatic hydrogens for substitution with a diverse set of functional groups (F, Cl, Br, I, -CN, $-NH_2$, -OH, $-OCH_3$, $-CF_3$). To ensure high-fidelity 3D geometries, all analogs underwent a rigorous conformational search using the Experimental-Torsion Knowledge Distance Geometry (ETKDGv3) algorithm. The resulting conformers were energy-minimized using the MMFF94s force field with a convergence threshold achieved within a maximum of 2,000 iterations.

Molecular Docking and CNN-Scoring. Molecular docking was executed using the GNINA engine (v1.0.3), which integrates a deep learning convolutional neural network (CNN) scoring function. Docking parameters were set to an exhaustiveness of 16 and a CNN-scoring resolution of 0.5 Å. The binding poses were evaluated based on the CNN-affinity score (expressed in pK_d units and converted to kcal/mol) and the CNN-pose score. Ligand efficiency (LE) and Lipophilic Ligand Efficiency (LLE) were calculated as secondary metrics for lead prioritization.

Supporting Information

Supporting Information File 1:
File Name: KRAS_G12D_Library_Metrics.xlsx
File Format: MS Excel
Title: Physicochemical Descriptors and Binding Metrics for the 100 Adenosine Analogs.
Description: This file includes SMILES strings, Binding Affinities (kcal/mol), Ligand Efficiency (LE), LLE, Molecular Weight, and TPSA for all generated candidates.

Supporting Information File 2:
File Name: Libreria_Optimizada_3D.sdf

File Format: SDF (Structure Data File)
Title: Optimized 3D Structures and Conformations.
Description: Coordinate file containing the 3D-minimized geometries and metadata for the entire chemical ensemble.

Acknowledgements
The author would like to thanks to the open-source community behind RDKit and GNINA for facilitating advanced drug discovery tools.

Funding
N/A

References

1. Prior, I. A.; Hood, F. E.; Kendrick-Jones, J. The frequency of Ras mutations in cancer. *Cancer Res.* **2020**, *80*, 2969–2974.
2. Simanshu, D. K.; Nissley, D. V.; McCormick, F. RAS proteins and their regulators in human disease. *Cell* **2017**, *170*, 17–33.
3. Hunter, J. C.; Gentry, J. P.; Ganesan, S., et al. Biochemical and structural analysis of common cancer-associated KRAS mutations. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 4282–4287.
4. Waters, A. M.; Der, C. J. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. *Cold Spring Harb. Perspect. Med.* **2018**, *8*, a031435.
5. Hallin, J.; Engstrom, L. D.; Hargis, L., et al. The KRASG12D inhibitor MRTX1133 targets KRASG12D in a noncovalent manner. *Cancer Discov.* **2022**, *12*, 1502–1517.
6. Wang, X.; Allen, S.; Blake, J. F., et al. Identification of MRTX1133, a Noncovalent, Potent, and Selective KRAS G12D Inhibitor. *J. Med. Chem.* **2022**, *65*, 3123–3133.
7. Ostrem, J. M.; Peters, U.; Sos, M. L., et al. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **2013**, *503*, 548–551.
8. Rosemeyer, H. The Chemodiversity of Purine as a Constituent of Natural Products. *Chem. Biodivers.* **2004**, *1*, 361–401.
9. Agrofoglio, L. A.; Gillaizeau, I.; Charluet, Y. Aza-and carba-nucleosides: structure and synthesis. *Tetrahedron* **2003**, *59*, 2841–2859.
10. Western, E. C.; Shaughnessy, K. H. Efficient, One-Step Suzuki Arylation of Unprotected Halonucleosides in Aqueous Media. *Org. Lett.* **2003**, *5*, 4545–4548.
11. Hocek, M. Synthesis of Purine-Based C-Nucleosides. *J. Org. Chem.* **2003**, *68*, 5773–5776.
12. Sunseri, J.; Koes, D. R. Virtual Screening with Gnina 1.0. *J. Chem. Inf. Model.* **2021**, *61*, 4849–4856.
13. Shaughnessy, K. H.; Booth, R. S. Pd-catalyzed cross-coupling reactions in aqueous media. *Chem. Soc. Rev.* **2006**, *35*, 199–213.
14. Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430–431.
15. Shultz, M. D. Two decades of anti-targets, ligand efficiency, and lipophilic efficiency. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5980–5991.

16. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.
17. Veber, D. F.; Johnson, S. R.; Cheng, H. Y., et al. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623.
18. Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **2007**, *6*, 881–890.
19. Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* **2017**, *57*, 942–957.
20. Hann, M. M. Molecular obesity, potency and other addictions in drug discovery. *MedChemComm* **2011**, *2*, 349–355.
21. Hann, M. M.; Keserü, G. M. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nat. Rev. Drug Discov.* **2012**, *11*, 355–365.
22. Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach. *ACS Chem. Neurosci.* **2010**, *1*, 435–449.
23. Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* **2005**, *10*, 464–469.
24. Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular weight, lipophilicity, hydrogen bond counts, and molar refractivity. *J. Med. Chem.* **1996**, *39*, 2887–2893.
25. Mao, Z.; Xiao, P.; Shen, P. Identification of a specific cholesterol binding site in the KRAS–PDEδ complex. *Chem. Sci.* **2022**, *13*, 8758–8766.
26. Patgiri, A.; Jernigan, F. E.; Arora, P. S. Rational design of a helix-mimetic inhibitor. *Nat. Chem. Biol.* **2011**, *7*, 585–587.
27. Pai, E. F.; Krengel, U.; Petsko, G. A., et al. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *Nature* **1990**, *341*, 209–214.
28. Ryckmans, T.; Edwards, M. P.; Horne, V. A., et al. Rapid assessment of a novel series of selective COX-2 inhibitors using a parameter optimization approach. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 4857–4859.