



OPEN ACCESS

EDITED BY

Bibekanand Mallick,
National Institute of Technology Rourkela,
India

REVIEWED BY

Luca Zanella,
Columbia University Irving Medical
Center, United States
Jing Wang,
Chinese Academy of Sciences (CAS),
China

*CORRESPONDENCE

Abdullah Kahraman,
✉ abdullah.kahraman@fhnw.ch

RECEIVED 23 December 2025

REVISED 16 January 2026

ACCEPTED 29 January 2026

PUBLISHED 27 February 2026

CITATION

Steuer J and Kahraman A (2026)
Unlocking the undruggable spliceosome:
generative AI and structural dynamics in
cancer therapy.
Front. Cell Dev. Biol. 14:1774239.
doi: 10.3389/fcell.2026.1774239

COPYRIGHT

© 2026 Steuer and Kahraman. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Unlocking the undruggable spliceosome: generative AI and structural dynamics in cancer therapy

Jakob Steuer^{1,2} and Abdullah Kahraman^{1,2*}

¹Data Science in Life Sciences Group, Institute for Chemistry and Bioanalytics, School of Life Sciences, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland, ²Swiss Institute of Bioinformatics, Amphipôle, Quartier UNIL-Sorge, Lausanne, Switzerland

The spliceosome is a dynamic molecular machine essential for transcriptome diversity, yet its complexity creates specific vulnerabilities in cancer. Recurrent somatic mutations in core factors, particularly SF3B1, U2AF1, and SRSF2, drive malignancies by altering splice-site recognition. Such structural perturbations do not merely drive oncogenesis but manifest as distinctive molecular signatures that can serve as potent diagnostic and prognostic biomarkers. However, therapeutic exploitation of these defects remains challenging. This review argues that unlocking the spliceosome requires a shift from static cryo-EM snapshots to dynamic structural ensembles. We explore how physics-based molecular simulation and enhanced sampling methods are merging with generative Artificial Intelligence to identify intermediate states, map cryptic allosteric pockets and target intrinsically disordered regions. Translating these mechanistic insights into the clinic, we evaluate the next-generation of therapeutic strategies, ranging from novel molecular biomarkers to rationally designed allosteric modulators and synthetic lethality. Finally, we discuss how deciphering these altered structural dynamics can guide the identification of splicing-derived neoantigens and biomarkers, establishing a roadmap for precision immunotherapy.

KEYWORDS

biomarker, cancer, generative AI, molecular dynamics, neoantigens, spliceosome

1 Introduction

The faithful expression of the eukaryotic genome is dependent on the spliceosome, a megadalton ribonucleoprotein (RNP) complex that orchestrates the removal of introns and the ligation of exons from precursor messenger RNA (pre-mRNA). This process is not merely a housekeeping function, but rather a fundamental engine of proteomic diversity. Through alternative splicing, a limited number of genes generate a vast repertoire of protein isoforms with distinct functions. These processes drive cellular differentiation, tissue homeostasis, and organismal complexity (Chen and Manley, 2009; Wang et al., 2015). In humans, more than 95% of multi-exon genes are reported to undergo alternative splicing, a flexibility that is essential for life but also represents a critical vulnerability (Lu et al., 2012; Jiang and Chen, 2021). The spliceosome, a complex intron-retaining ribonucleoprotein, is made up of five small nuclear RNAs (snRNAs) and hundreds of proteins. This intricate assembly necessitates a highly coordinated process of assembly and disassembly on each pre-mRNA substrate.

Within the context of cancer, this machinery is frequently co-opted. Somatic mutations in genes encoding core splicing factors, most notably *SF3B1*, *U2AF1*, and *SRSF2*, have emerged as cardinal drivers of hematologic malignancies such as myelodysplastic syndromes (MDS) (Yoshida et al., 2011) and chronic lymphocytic leukemia (CLL) (Quesada et al., 2012), as well as solid tumors like uveal melanoma and breast cancer (Harbour et al., 2013; Ellis et al., 2012). In contrast to typical loss-of-function mutations observed in tumor suppressor genes, these are heterozygous, change-of-function mutations that fundamentally alter the structural dynamics of spliceosome assembly (Zhang et al., 2021). The process does not result in the destruction of the splicing machinery. Instead, it involves the reprogramming of its thermodynamic preferences, thereby redirecting it to recognize aberrant splice sites (Zhang et al., 2024). This generates tumor-specific isoforms that can drive oncogenesis, confer drug resistance, or create potentially immunogenic neoantigens. Integrating structural dynamics with large-scale transcriptomics (multi-omics) allows for the identification of isoform-level biomarkers. These signatures can stratify patients into risk groups, serving as prognostic indicators for disease progression in MDS and CLL (Tseng and Obeng, 2024; Qiu et al., 2016; Zhang et al., 2017).

For decades, the spliceosome was considered an “undruggable” target because of its sheer size, lack of deep hydrophobic pockets in its static states, selectivity challenges and extraordinary dynamic nature. The splicing cycle involves a continuous, ATP-driven remodeling of the U1, U2, U4, U5, and U6 snRNPs (Wahl et al., 2009). While the “resolution revolution” in cryogenic electron microscopy (cryo-EM) has provided atomic-level maps of the spliceosome’s discrete conformational states, these structures historically represent snapshots of deep energy minima (Fica and Nagai, 2017; Tholen and Galej, 2022). They often fail to capture the “breathing” motions, intrinsic disorder, and transient transition states where therapeutic opportunities frequently lie. A convergence of technological breakthroughs is reshaping this landscape. The integration of computational biophysics, particularly Molecular Dynamics (MD) simulations, with generative Artificial Intelligence (AI) is enabling a shift from static structural snapshots to rich dynamic ensembles. These approaches make it possible to characterize the fluctuations of splicing factors, reveal transient “cryptic” allosteric pockets (Vajda et al., 2018) exposed during the splicing cycle, and computationally design small molecules and *de novo* protein binders that engage intrinsically disordered regions (IDRs), which have traditionally been deemed intractable (Sun et al., 2025).

The mutational landscape (Zhang et al., 2024), cryo-EM based structural foundations (Vorländer et al., 2022; Tholen and Galej, 2022), and clinical potential (Eymin, 2021; Anczukow et al., 2024; Lv et al., 2025) of the spliceosome have been extensively reviewed. More recently, the field has begun to appreciate the machinery’s fluid nature, and how molecular simulations complement experimental data to decode spliceosome dynamics (Borišek et al., 2023; 2021; Pokorná et al., 2025). This review builds upon these mechanistic foundations but pivots toward the therapeutic application of these dynamics using emerging technologies. We move beyond the characterization of equilibrium fluctuations to explore how Generative AI and biomolecular language models are reshaping

the design of binders for transient pockets and intrinsically disordered regions. Finally, we extend this structural perspective into the clinical realm, linking dynamic conformational defects directly to synthetic lethal vulnerabilities and the systematic discovery of splicing-derived neoantigens, proposing a forward-looking roadmap for next-generation precision medicines.

2 Dynamics of the canonical splicing cycle

Understanding the pathological deviations observed in cancer requires a general grasp of the healthy splicing cycle (Figure 1A). The spliceosome operates as a metallo-ribozyme: a catalytic RNA core coordinated by metal ions, encased within a protein scaffold that ensures fidelity (Fica et al., 2013). The splicing cycle is driven by ATP-dependent DExD/H-box RNA helicases, which power the conformational transitions required for assembly, activation, catalysis, and disassembly (Dörner and Hondele, 2024).

2.1 Early recognition and the E-Complex

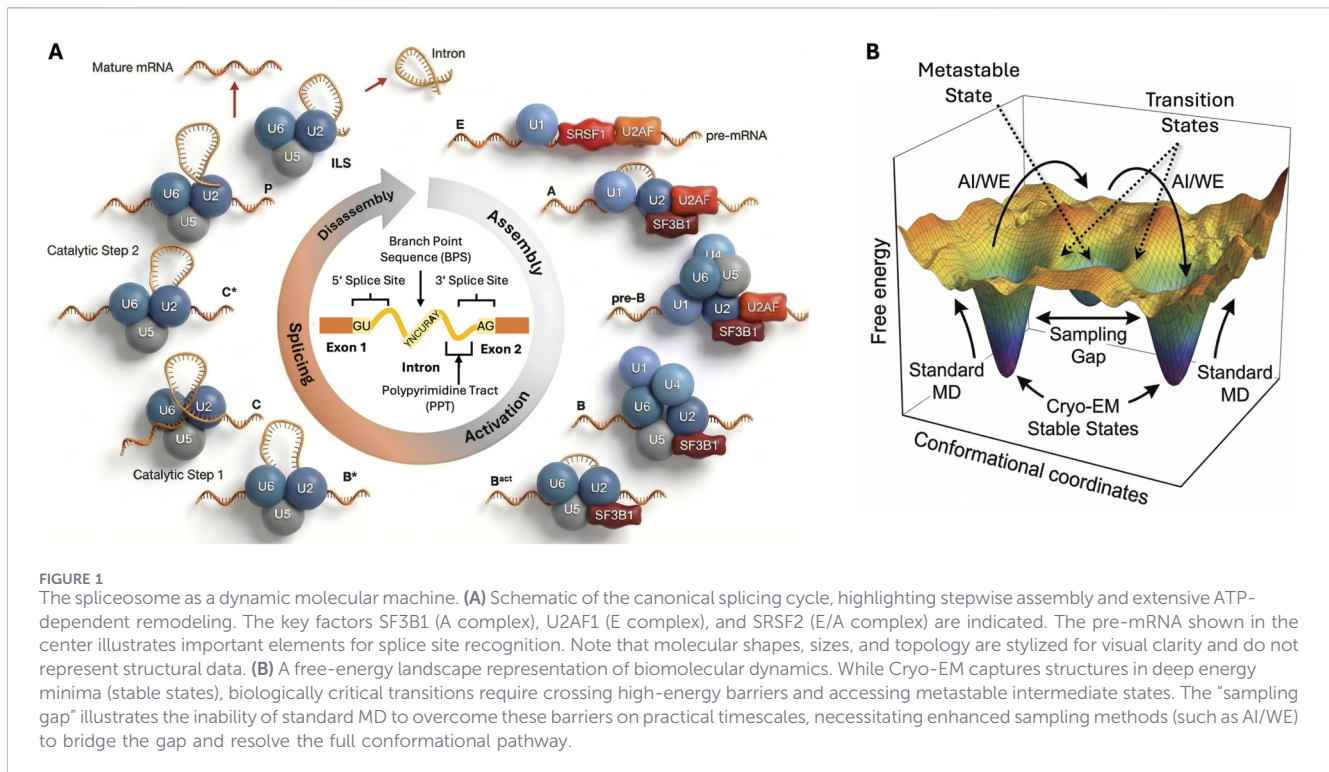
Assembly initiates with the E complex, a cross-intron scaffold that establishes the molecular foundation for splice-site selection. U1 snRNP recognizes the 5' splice site (5'SS), while the 3' region is defined by the U2 auxiliary factor (U2AF) heterodimer. The large subunit, U2AF2, uses a flexible linker to bind the polypyrimidine tract (PPT), providing a modular anchor Sickmier et al. (2006). Concurrently, the small subunit U2AF1 secures the intron-exon boundary by specifically recognizing the conserved AG dinucleotide at the 3' splice site (Wu et al., 1999).

2.2 Checkpoint control in the A-Complex

The transition to the A complex represents a decisive commitment step. Here, U2 snRNP base-pairs with the branch point sequence (BPS) to form the Branch Point Helix. This interaction bulges the branch adenosine, exposing it for future catalysis. The architecture is stabilized by the SF3b subcomplex, specifically SF3B1, which clamps around the helix using an induced-fit mechanism to position the nucleophile (Cretu et al., 2016). This stage acts as a critical fidelity checkpoint: the helicase DDX46 (Prp5) utilizes ATP hydrolysis to test the stability of the U2-BPS interaction, rejecting suboptimal sites via kinetic proofreading (Xu and Query, 2007).

2.3 Catalytic activation and intrinsic disorder

Subsequent recruitment of the U4/U6.U5 tri-snRNP triggers extensive remodeling by the helicase Brr2, which unwinds the U4/U6 duplex to release U6 for catalysis (Ragunathan and Guthrie, 1998). The resulting catalytic core functions as a metallo-ribozyme, mediating transesterification via RNA-coordinated divalent metal ions (Mg^{2+}). However, final structural competence requires the DExD/H-box helicase Prp2 (DHX16). Prp2 destabilizes the SF3b subcomplex, effectively displacing it to expose the branch adenosine to the active site for the first reaction step (Ohrt et al., 2012; Wan et al., 2016).



Beyond these defined structures, the spliceosome relies on a high degree of intrinsic disorder. Proteins such as SRSF1 and the linkers between U2AF subunits utilize disordered regions to facilitate "fly-casting" mechanisms, effectively increasing the capture radius for molecular partners (Mackereth et al., 2011). While these fluid segments drive liquid-liquid phase separation (LLPS) and the formation of nuclear speckles (Kato et al., 2012), their conformational heterogeneity poses significant challenges for high-resolution modeling, representing a major frontier for structure-based drug design.

3 Pathogenic remodeling: how mutations rewire the spliceosome

Cancer-associated somatic mutations in the spliceosome are not random loss-of-function events but specific, positively selected lesions that reprogram the structural logic of RNA recognition (Yoshida et al., 2011; Dvinge et al., 2016). In the following, we will focus on recurrent mutations in core factors SF3B1, U2AF1, and SRSF2 to illustrate how discrete structural perturbations can drive transcriptome-wide oncogenic shifts, making them promising targets in therapeutic strategies.

3.1 SF3B1: the gatekeeper of branch point fidelity

The most frequent splicing lesions occur in *SF3B1*, with the K700E substitution in the HEAT repeat domain serving as the prototype in myelodysplastic syndromes and uveal melanoma (Papaemmanuil et al., 2011; Harbour et al., 2010). Structurally, the replacement of a positively charged lysine with a negatively

charged glutamic acid disrupts the local electrostatics near the branch point helix (Cretu et al., 2016) (Figure 2A). Critically, this mutation uncouples SF3B1 from the cofactor SUGP1, which is normally required to recruit the proofreading helicase DHX15. The failure to recruit DHX15 effectively bypasses the kinetic proofreading checkpoint described above, allowing the spliceosome to engage aberrant, upstream branch points (Zhang et al., 2022; Xing et al., 2025). This geometric shift forces the selection of cryptic 3' splice sites, generating pseudo-exonic inclusions in tumor suppressors such as *BRD9*, which subsequently undergo nonsense-mediated decay (Inoue et al., 2019). The detection of these cryptic 3' splice sites in peripheral blood or bone marrow represents a potential high-sensitivity diagnostic biomarker for SF3B1-mutant clones, preceding clinical symptoms of hematologic transformation (Bak-Gordon and Manley, 2025; Dolatshad et al., 2016; Malcovati et al., 2020).

3.2 U2AF1: altering zinc finger specificity

Mutations in *U2AF1* fundamentally alter the recognition of the 3' splice site AG dinucleotide through steric mechanisms (Yoshida et al., 2020; 2013). The recurrent S34F and S34Y mutations involve the substitution of a serine residue in the first zinc knuckle with a bulky aromatic side chain (phenylalanine or tyrosine, respectively). Structural analysis of the S34Y variant reveals that this substitution introduces a steric clash that reshapes the RNA-binding pocket, reducing the protein's structural flexibility (Figure 2B). While the wild-type protein efficiently recognizes both pyrimidines at the -3 position (CAG or UAG), the mutant form exhibits a strict preference for cytosine while disfavoring uridine (Okeyo-Owuor et al., 2015; Ilagan et al., 2015). This altered specificity creates a transcriptome-wide filter: exons preceded by a cytosine are

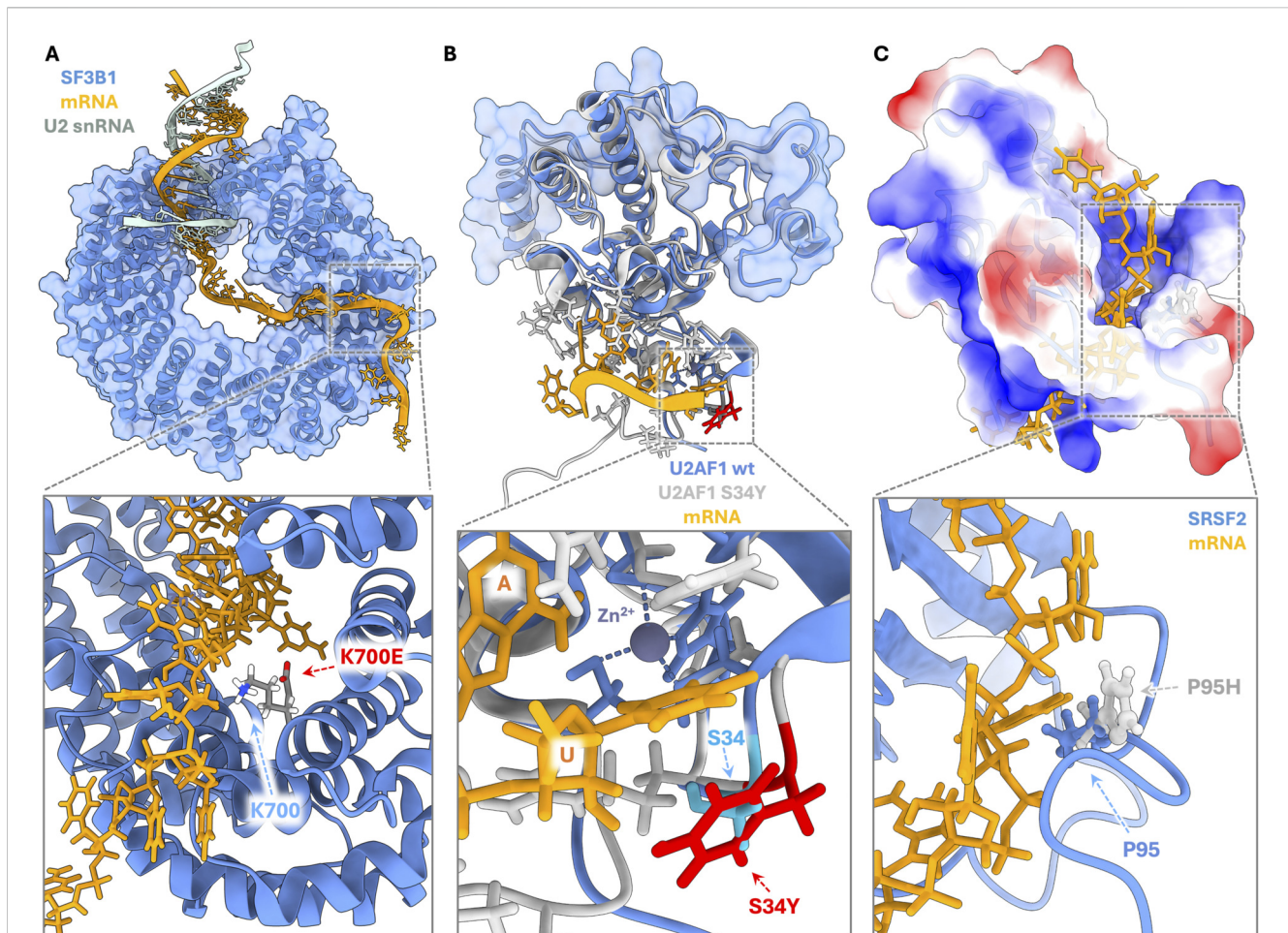


FIGURE 2

Atomic-level rewiring of RNA recognition by oncogenic mutations. (A) SF3B1 K700E. The heat-repeat domain of SF3B1 acts as a scaffold for the branch point sequence (BPS). The K700E substitution (modeled here on PDB ID 5Z56 (Zhang et al., 2018)) introduces a negative charge into the RNA-binding tunnel. This alters the local electrostatic environment, destabilizing the interaction with the canonical branch point adenosine. (B) U2AF1 S34F. An overlay of the wild-type U2AF1 zinc knuckle (blue and cyan, PDB ID 4YH8 (Yoshida et al., 2015)) and the S34Y mutant complexed with RNA (grey and red, PDB ID 7C08 (Yoshida et al., 2020)). The S34Y mutation serves as a structural surrogate for the clinical S34F variant. The bulky aromatic side chain creates a steric clash with uridine at the -3 splice site position, driving a shift in specificity toward cytosine to relieve steric hindrance. (C) SRSF2 P95H. The Pro95 residue is critical for the orientation of the linker region relative to the RNA Recognition Motif (RRM) (PDB ID 2LEB (Daubner et al., 2012)). The P95H substitution alters the conformational equilibrium of the protein, changing its affinity to favor C-rich over G-rich Exonic Splicing Enhancer (ESE) motifs. Electrostatic surface potential illustrates the structural interface for mRNA recognition. Methodological Note: For panels (A) and (C), where specific mutant crystal structures were unavailable, mutant side chains were modeled using the Dunbrack rotamer library within UCSF ChimeraX (Meng et al., 2023).

preferentially retained, while those with a uridine are skipped, driving the mis-splicing of oncogenes like *H2AFY* and *IRAK4* (Kim et al., 2021; Smith et al., 2019). Against this backdrop, *U2AF1* S34 mutations are now incorporated into molecular risk models as adverse prognostic biomarkers in myeloid neoplasms, and their associated splice isoform signatures (for example, *IRAK4-L*) are under active investigation as companion diagnostics for emerging *IRAK4* and spliceosome-targeted therapies in precision oncology (Garcia-Manero et al., 2024; Badar et al., 2023; Li et al., 2020).

3.3 SRSF2: modulating exonic splicing enhancers

In contrast to intronic recognition defects discussed before, SRSF2 P95 hotspot mutations (primarily P95H, P95L, and P95R)

remodel the recognition of exonic splicing enhancers (ESEs) (Daubner et al., 2012; Kim et al., 2015). P95 resides in the linker region between the RNA recognition motif (RRM) and the RS domain, where mutations induce conformational changes affecting the RRM termini (Figure 2C). This structural alteration shifts SRSF2's RNA-binding specificity: whereas wild-type SRSF2 recognizes the degenerate SSNG motif (S=C or G) with similar affinity for both CCNG and GGNG variants, mutant SRSF2 exhibits enhanced binding to CCNG sequences and reduced affinity for GGNG motifs. This specificity switch drives widespread alterations in cassette exon inclusion and exclusion that can lead to nonsense-mediated decay, production of aberrant protein isoforms with altered function, and impaired hematopoietic differentiation (Liu et al., 2024; Rahman et al., 2020; Kim et al., 2015; Liang et al., 2018). At the clinical level, SRSF2 P95 mutations define a characteristic subset of myeloid

neoplasms associated with adverse prognosis and are being incorporated into molecular risk stratification models (Hoff et al., 2024). Beyond prognostication, the distinct splicing signatures driven by these mutations are under active investigation as biomarkers for precision oncology, highlighting synthetic lethal vulnerabilities that currently remain in preclinical development (Su et al., 2024; Eldfors et al., 2024).

4 The sampling gap: challenges in simulating large-scale dynamics

Molecular dynamics (MD) has been an important tool for generating atomistic hypotheses about spliceosome structure-function relationships, refining cryo-EM models, probing local flexibility, and rationalizing the impact of mutations. Pokorná et al. provide an excellent review of the major MD studies on spliceosomal components (Pokorná et al., 2025). However, using only classical equilibrium MD as a primary engine for spliceosome drug discovery exposes several fundamental limitations, as the spliceosome pushes the boundaries of system size, timescale, and force-field accuracy (Baltrukevich and Bartos, 2023; Muscat et al., 2024; Migens et al., 2025).

The most fundamental constraint is the disparity between accessible simulation times and biological timescales (Figure 1B). Even on modern GPU clusters and specialized hardware, routine all-atom MD for systems the size of spliceosomal subcomplexes typically reaches microseconds at best. Achieving tens to hundreds of microseconds is possible but extremely costly, and millisecond trajectories remain exceptional rather than routine (Shaw et al., 2009; Gapsys et al., 2024). In contrast, functionally important spliceosomal transitions, such as the large-scale remodeling of snRNPs (e.g., B → B* activation), rearrangements of HEAT-repeat domains, or the opening of deep, transient pockets, occur on millisecond to second timescales. These motions often involve crossing substantial free-energy barriers and visiting sparsely populated intermediate states. This 3- to 6-order-of-magnitude gap means that straightforward equilibrium MD will overwhelmingly sample fluctuations near the starting structure (Hollingsworth and Dror, 2018).

Classical force fields such as AMBER and CHARMM have been extensively optimized for proteins and, more recently, substantially improved for RNA and protein-RNA complexes (Mlýnský et al., 2025). Nonetheless, spliceosomal systems present particularly stringent tests. RNA's densely charged phosphate backbone and complex stacking/branching modes make its behavior highly sensitive to the treatment of long-range electrostatics, water models, and ion parameters. Residual inaccuracies can artificially weaken or over-stabilize RNA-protein interfaces, distort backbone conformations, or make folded RNA elements appear either too rigid or too labile compared to experiment (Mlýnský et al., 2025; Baltrukevich and Bartos, 2023; Šponer et al., 2018). Additionally, standard non-polarizable models fail to capture the complex coordination geometry and charge transfer of catalytic Mg²⁺ and K⁺ ions, often necessitating expensive QM/MM (Quantum Mechanics/Molecular Mechanics) approaches for accurate active site representation (Grotz et al., 2021; Šponer et al., 2018). And while explicit solvent is required to capture essential water-mediated

interactions, its computational cost limits the exploration of rare transitions, a bottleneck that implicit solvent models cannot reliably address without distorting the delicate balance of protein-RNA forces (Chen et al., 2023). Sampling is further hindered by the rugged energy landscape. Because cryo-EM structures represent deep free-energy minima, unbiased MD rarely accesses higher-energy intermediates where allosteric pathways and cryptic pockets often reside, effectively inferring the plot of a movie from its opening frame (Vant et al., 2022; Verkhivker et al., 2023).

5 Enhanced sampling of rare events

To address the sampling bottlenecks of classical MD, a range of enhanced sampling strategies and statistical frameworks have been developed. These methods deliberately bias dynamics, restructure sampling protocols, or reduce system resolution to recover statistics on rare events and massive conformational changes standard MD cannot access. Reviews of enhanced sampling techniques for biomolecular simulations provide detailed guidance on methodology selection and implementation across diverse timescale regimes (Mehdi et al., 2024; Hooten et al., 2025; Yang et al., 2019). In the following, we focus on promising techniques for studying spliceosomal dynamics and cryptic binding site discovery on splicing factors.

5.1 Cryptic pockets and binding thermodynamics

Recent advances in identifying cryptic pockets and exploiting allosteric mechanisms have transformed the drug discovery landscape, enabling the targeting of previously inaccessible protein sites (Bemelmans et al., 2025; Mukhopadhyay et al., 2025). Mixed-solvent molecular dynamics (MSMD) simulations augment conventional MD by introducing small organic probes (e.g., benzene, isopropanol) that can identify and map these cryptic binding sites. (Mukhopadhyay and Chakrabarty, 2025; Vithani et al., 2024; Koseki et al., 2025; Chan et al., 2023; Wakefield et al., 2022; Smith and Carlson, 2021; Kuzmanic et al., 2020). These probes compete with water, clustering in thermodynamically favorable "hotspots" that reveal ligandable sites on flexible surfaces. Recent advances in topological data analysis combined with MSMD (Koseki et al., 2025) have significantly improved cryptic site prediction accuracy by assessing protein conformational variability during probe clustering, substantially outperforming recent machine-learning approaches. While MSMD identifies where ligands bind, alchemical methods such as Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) quantify how strong that binding is (York, 2023). By physically transforming one ligand into another along a non-physical pathway, these methods calculate relative binding free energies with high accuracy (Ross et al., 2023; Sampson et al., 2024).

5.2 Kinetics and rare transition pathways

The Weighted Ensemble (WE) method accelerates rare events without biasing the underlying energy landscape (Chong and Zuckerman, 2025; Aristoff et al., 2023; Hellemann and Durrant,

2023). It partitions configuration space along a progress coordinate, splitting trajectories in under-sampled regions and merging them in over-sampled ones. Complementing WE, Markov State Models (MSMs) provide a framework to analyze aggregate simulation data (Kononov et al., 2021; Bose et al., 2025; Liu et al., 2025). By discretizing high-dimensional conformational space into countable states and estimating transition probabilities between them, MSMs reconstruct long-timescale kinetics from many short, parallel MD trajectories. Recent application to RNA-ligand binding demonstrated MSM's power for characterizing complex kinetic pathways in neomycin-riboswitch binding (Chyzy et al., 2024). For spliceosomal proteins, MSMs could similarly map the landscape of branch point adenosine and polypyrimidine tract recognition by SF3B1 and U2AF2, revealing metastable intermediates that may represent kinetic bottlenecks or allosteric vulnerabilities for therapeutic targeting.

5.3 Large-scale dynamics and sampling

The megadalton size of the spliceosome often precludes all-atom sampling of global rearrangements. Coarse-Grained (CG) MD models (e.g., Martini, Gō -models) overcome this by grouping atoms into single interaction beads, reducing degrees of freedom and allowing larger time-steps (Souza et al., 2021; Yangaliev and Ozkan, 2025; Takada, 2019). CG is essential for simulating the massive domain reorganizations required during the spliceosome cycle, such as the transition from the B to the B^{act} complex, or the phase-separation behavior of intrinsically disordered regions in SR proteins (Zhan et al., 2024; Von Bülow et al., 2025).

For systems where atomistic detail is required but barriers are high, methods like Metadynamics apply history-dependent potentials to specific variables to force transitions (Laio and Parrinello, 2002). This forces the system out of local minima, allowing the calculation of free-energy landscapes for specific reaction coordinates, such as the bending of the SF3B1 superhelix or the mechanical opening of the branch point pocket (Rakesh et al., 2016; Zhan et al., 2024). Alternatively, Replica-exchange MD (REMD) and related schemes facilitate barrier crossing by running multiple simulations at different temperatures (or Hamiltonians) and periodically swap configurations (Zhu et al., 2025). While applying REMD to entire spliceosomal assemblies is typically prohibitive, targeted use on isolated domains or intrinsically disordered regions can improve exploration of alternative conformations.

6 From static structure to generative design

We are in the midst of a paradigm shift in structural biology driven by deep learning and generative AI (Malhotra et al., 2025). The field is moving beyond predicting a single “best” structure toward modeling thermodynamic ensembles, interfaces, and designing new biomolecules *de novo* (Cui et al., 2025; Schapira et al., 2024). Rather than explicitly integrating equations of motion as in Molecular Dynamics (MD), these models learn statistical regularities in sequence, structure, and interaction data. In favorable cases, they can sidestep the sampling bottlenecks that

limit classical simulations, offering a new toolkit for targeting the spliceosome (Desai et al., 2024).

6.1 Foundation models for structural prediction

AlphaFold2 (AF2) established that protein structures could be predicted with near-experimental accuracy for many monomers (Jumper et al., 2021). Its successor, AlphaFold3 (AF3), extends this concept with a unified diffusion-based architecture capable of modeling protein-protein, protein-ligand, and protein-nucleic acid complexes within a single framework (Abramson et al., 2024). For spliceosomal biology, such models may provide a way to generate atomistic hypotheses for mutant splicing factors bound to RNA elements or small-molecule inhibitors that are experimentally challenging to crystallize.

In parallel, a growing ecosystem of open-weights foundation models has emerged, democratizing access to complex prediction. RoseTTAFold All-Atom (RFAA) pioneered the generalized modeling of protein-nucleic acid-small molecule complexes, offering an open-source alternative to proprietary systems (Krishna et al., 2024). Building on this, Boltz-2 represents a significant leap forward, predicting not only 3D structures of mixed-modality complexes (proteins, RNA, DNA, small molecules) but also binding affinities, achieving accuracy competitive with or superior to AF3 in recent benchmarks (Passaro et al., 2025; Xu et al., 2025). Similarly, CHai-1 offers a robust alternative for predicting protein-nucleic acid interfaces (Chai et al., 2024).

Despite their power, it is important to note that these models are not physics engines. They can “hallucinate” plausible-looking interactions for physically impossible contacts (e.g., steric clashes in novel drug scaffolds) or predict rigid structures for intrinsically disordered regions (IDRs) that should remain unstructured (Krokidis et al., 2025; Fang et al., 2025). Thus, the outputs should be treated as high-confidence hypotheses rather than ground truth (Schapira et al., 2024).

6.2 From single structures to dynamic ensembles

While powerful, the single-structure output of these recently developed foundation models reflects an inherent bias toward one thermodynamically stable conformation, usually the dominant state in experimental training data. However, the spliceosome is a heat-driven molecular machine that relies on distinct, often transient, conformations to function. Capturing this heterogeneity is essential for understanding mechanistic transitions, such as the destabilization of the B^{act} complex prior to catalysis or the conformational release of SF3B1, states that are often invisible to static crystallography but could be accessible to generative ensembles.

New algorithms are emerging as “generative MD surrogates,” sampling diverse conformations at a fraction of the computational cost of physical simulations (Cui et al., 2025; Aranganathan et al., 2025). BioEmu uses a diffusion-based framework to estimate equilibrium distributions directly from sequence, effectively capturing the Boltzmann distribution of states (Lewis et al.,

2025). Similarly, AlphaFlow/ESMFlow adapts the AF2 architecture using flow matching to interpolate between noise and diverse structural states, providing a “pseudo-trajectory” of conformational flexibility (Jing et al., 2024). While these models sample equilibrium states orders of magnitude faster than MD, they forfeit temporal coherence, rendering them unable to predict transition kinetics or the causal pathways of conformational switching.

Complementary strategies steer existing predictors toward higher diversity. AF-Cluster systematically clusters or subsamples the Multiple Sequence Alignment (MSA) to reduce evolutionary constraints (Wayment-Steele et al., 2024; Sala et al., 2023). In contrast, Entropy Guidance (EGF) utilizes the full MSA but modifies AlphaFold2’s intermediate hidden states via an auxiliary loss function to disrupt the model’s bias (Wu and Feng, 2025). While these heuristics effectively identify alternative minima, they inherently lack thermodynamic calibration. AlphaFold2-RAVE (Vani et al., 2023) bridges this gap by using these diverse predictions to seed physics-based enhanced sampling, successfully recovering Boltzmann-weighted ensembles. However, strategies that rely on MSA subsampling (including AF-Cluster and RAVE’s initialization) remain limited by alignment depth, hindering application to rapidly evolving variants where evolutionary data is sparse.

6.3 Inverse design of binders and language models

Beyond understanding existing structures, generative AI is now tackling the “inverse design” problem: proposing sequences that fold into specific structures or programming RNA sequences to control splicing outcomes.

ProteinMPNN and LigandMPNN have become standard tools for sequence recovery, rapidly designing sequences for fixed backbones or small-molecule binders (Dauparas et al., 2022; 2025). However, the spliceosome requires targeting complex Ribonucleoprotein (RNP) interfaces. The recently introduced RFDpoly (RoseTTAFold Diffusion for Biopolymers) (Favor et al., 2025) addresses this gap, building on the RFdiffusion model to enable the simultaneous hallucination of nucleic acid backbones and protein binders (Watson et al., 2023). This allows for the *de novo* design of proteins that can lock specific RNA conformations, or synthetic RNAs that mimic U-snRNA structural motifs. A unique challenge in splicing is targeting Intrinsically Disordered Regions (IDRs), such as the RS domains of SR proteins. Here, generative approaches must be adapted to design peptide binders that remain flexible yet possess high affinity, decoupled from rigid folding constraints.

Newer pipelines like BindCraft integrate AlphaFold-Multimer weights with ProteinMPNN and physics-based filters to perform “one-shot” design of high-affinity binders Pacesa et al. (2025). This composite approach effectively targets challenging interfaces that often evade traditional docking. BoltzGen extends this frontier further, introducing a unified all-atom generative model that enables the co-design of binders for diverse modalities, including proteins, nucleic acids, and small molecules, with explicit structure-based reasoning (Stark et al., 2025). A key challenge are “adversarial” sequences, that have high-confidence structural prediction (e.g.,

high pLDDT or pAE scores) despite being biologically invalid, physically unstable, aggregation-prone, or unfolded in reality. This has made rigorous post-design filtration using methods like BindEnergyCraft for physics-based rescoring or orthogonal co-folding assays an essential step to separate viable candidates from these high-confidence false positives (Nori et al., 2025).

Parallel to structural design, Large Language Models (LLMs) are decoding the “grammar” of splicing regulation. While geometric models handle 3D coordinates, emerging RNA foundation models like RiNALMo (Penić et al., 2025), and ERNIE-RNA (Yin et al., 2025) leverage vast transcriptomic datasets to predict secondary structure and function directly from sequence, outperforming earlier models on unseen RNA families. For 3D design, gRNAd applies geometric deep learning to the inverse problem, conditioning on 3D backbones to generate sequences that adopt specific tertiary folds (Joshi et al., 2024). On the functional side, TrASPr (Transformer for Alternative Splicing Prediction) represents a specialized leap forward (Wu et al., 2025). Unlike general protein LLMs (e.g., ESM3), TrASPr is explicitly trained to predict tissue-specific splicing outcomes and can design RNA sequences to modulate inclusion levels of specific exons.

6.4 Validation of structural fidelity for reliable binder and biomarker prediction

As generative models for protein design and RNA processing proliferate, rigorous benchmarking is essential to distinguish genuine predictive power from hallucination. This validation is critical across all modalities discussed, from the *de novo* design of binders for disordered regions to the large language models (LLMs) predicting splicing outcomes. Recent benchmarking efforts have revealed critical limitations. Studies such as FoldBench and Runs N’ Poses have addressed the risk of data leakage, where structural redundancies between training and test sets inflate performance metrics (Xu et al., 2025; Škrinjar et al., 2025). By strictly segregating datasets based on sequence identity and structural similarity, these benchmarks assess true generalization capabilities. Results indicate that current performance relies heavily on memorization; accuracy declines sharply in out-of-distribution (OOD) scenarios, such as when models are tasked with predicting novel ligands or binding pockets dissimilar to those in the training set. Similarly, PoseBusters was developed to audit the physical viability of generated molecules by validating geometric accuracy and chemical plausibility, such as assessing bond lengths and flat angles. It demonstrated that even geometrically accurate models can produce chemically implausible structures with steric clashes (Buttenschoen et al., 2024). Furthermore, recent results from CASP16, a community experiment to advance methods of computing three-dimensional protein structure, underscored persistent challenges in complex prediction, particularly for antibody-antigen targets (Zhang et al., 2025). These limitations suggest that while AI can theoretically be used for spliceosome targeting or to predict events such as “splicing switch-points”, the structural transitions that lead to drug resistance, the current rate of hallucination and OOD failure poses a risk to both therapeutic design and biomarker discovery. Consequently, AI models function best as sophisticated filters to prioritize candidates for downstream

Free Energy Perturbation (FEP) and wet-lab validation, ensuring that AI-predicted candidates are biophysically sound before they enter the precision oncology pipeline.

7 Translating structural dynamics into clinical strategy

7.1 Direct modulation of the spliceosome core

Pharmacological strategies have shifted from broad-spectrum modulation toward precision targeting of the spliceosome machinery and its auxiliary factors (Lv et al., 2025; Tseng and Obeng, 2024; Anczukow et al., 2024). First-generation modulators, such as pladienolide B and its analogs (E7107, H3B-8800), exploit the structural plasticity of the SF3B1 subunit. Cryo-EM and structural studies reveal that these agents bind within the branch point adenosine (BPA) binding tunnel, locking the complex in an open, pre-catalytic conformation. MD simulations suggest that this binding event reduces the internal cross-correlation of SF3B1 HEAT repeats, effectively freezing the functional plasticity required to accommodate the branch point adenosine (Amorello et al., 2025; Corrionero et al., 2011; Yokoi et al., 2011). While these agents bind wild-type and mutant SF3B1 with equal affinity, they induce preferential lethality in mutant cells by exploiting the reduced tolerance for additional splicing stress (functional selectivity) (Seiler et al., 2018; Wheeler et al., 2024). However, the lack of thermodynamic selectivity resulted in a narrow therapeutic index (Jiang et al., 2023). E7107 was discontinued due to ocular toxicity, and H3B-8800 failed to achieve objective responses in myelodysplastic syndrome, hampered by on-target dose-limiting toxicities such as QTc prolongation (Hong et al., 2014; Steensma et al., 2019).

7.2 Programmable targeting *via* antisense oligonucleotides

While small molecules modulate the protein machinery, antisense oligonucleotides (ASOs) offer a programmable logic to target the RNA substrate directly (Takeshima, 2025; Havens and Hastings, 2016). ASOs are chemically modified nucleic acids that base-pair with pre-mRNA to sterically block regulatory motifs with high sequence specificity. However, the therapeutic utility of ASOs has historically been limited by the dynamic nature of RNA folding; target sequences are frequently occluded within stable secondary structures, rendering them thermodynamically inaccessible. Here, the integration of generative AI could solve the scalability bottleneck. New deep learning frameworks can now predict dynamic RNA accessibility profiles and optimize ASO sequences to outcompete internal RNA base-pairing (Wu et al., 2025; Lin et al., 2024; Kang et al., 2025).

7.3 Allosteric modulation and cryptic pockets

An emerging approach that circumvents these shortcomings is to develop allosteric inhibitors targeting cryptic pockets unique to

mutant conformations. Molecular dynamics simulations have successfully mapped dynamic drug binding landscapes, revealing transient, hydrophobic trenches between RNA recognition motifs (RRMs) in U2AF2 that are invisible to static crystallography (Borišek et al., 2020). Similarly, simulations have identified metastable pockets near the SF3B1 HEAT 5-7 repeats and the K700E mutation site, providing novel footholds for allosteric control (Rozza et al., 2023; Spinello et al., 2021). In contrast to pan-modulators, allosteric ligands aim to selectively bind the oncogenic variant, theoretically sparing essential wild-type splicing. Distinct from traditional inhibitors that block active sites, aryl sulfonamides function as “molecular glues,” acting as an interfacial wedge that reshapes the DCAF15 surface to create a high-affinity composite binding site for the α -helix 1 of the RBM39 RRM2 domain. Structural analysis confirms that the sulfonamide moiety mimics the side chain of an endogenous amino acid, allowing the complex to accommodate the RBM39 Gly268 residue, a site where any sterically larger mutations confer drug resistance (Han et al., 2017). This strategy causes a collapse in splicing efficiency and is currently under clinical evaluation (Bewersdorf et al., 2023).

7.4 Synthetic lethality

Chronic splicing stress creates downstream vulnerabilities in genome maintenance and protein homeostasis that can be exploited synthetically. Defective co-transcriptional splicing promotes R-loop accumulation and replication stress, rendering mutant cells critically dependent on ATR/Chk1 checkpoint signaling. Inhibitors of this axis show preferential toxicity in spliceosome-mutant models (Nguyen et al., 2018; Bland et al., 2023). This vulnerability extends to cells with cohesin mutations (e.g., *STAG2*), which lack the capacity to repair R-loop-induced damage, defining a convergent synthetic lethal axis (Wheeler et al., 2024). Transcriptome-wide mis-splicing burdens protein quality control systems. Beyond proteasome inhibition (Huang et al., 2020), recent studies highlight the Integrated Stress Response (ISR) as a key target. For example, *U2AF1*-driven mis-splicing sensitizes cells to ISR modulation (Jin et al., 2024). Furthermore, upstream snRNP biogenesis represents a critical bottleneck. PRMT5, which methylates Sm proteins to facilitate assembly, is essential for spliceosome-mutant cell survival. Inhibition of PRMT5 triggers a lethal collapse of snRNP availability, offering a distinct therapeutic avenue (Wu et al., 2024; Fong et al., 2019; Kohsaka et al., 2024). The clinical success of synthetic lethal approaches will depend on robust predictive biomarkers, such as defined splicing factor mutations and R-loop-associated replication-stress signatures, that can enrich for patients most likely to respond and thus operationalize precision oncology (Peng et al., 2025; Nguyen et al., 2018).

7.5 Splicing-derived neoantigens as immunotherapy targets

Beyond direct targeting of splicing factors, the aberrant transcripts they generate provide a second axis for therapy: tumor-specific neoantigens. By altering the choice of the splice site, splicing factor mutations create novel exon-exon junctions and retained introns that can be translated into peptide

sequences that are not present in the normal human proteome (Li et al., 2024; Kwok et al., 2025). The total splicing neoantigen burden, quantified through the integration of transcriptomic data and structural modeling, is emerging as a potent prognostic biomarker. High-burden tumors often exhibit enhanced immunogenicity with shared, targetable neoantigens, correlating with improved survival and superior response rates to immune checkpoint blockade (ICB) therapies in treated cohorts (Li et al., 2024).

Most neoantigen studies have focused on peptides arising from single nucleotide variants (SNVs), which typically introduce single amino acid substitutions. In contrast, aberrant splicing often causes frameshifts, premature termination, or inclusion of intronic sequences, leading to longer stretches of novel amino acids or entirely new C-terminal tails (Kwok et al., 2025; Smart et al., 2018). Importantly, these splicing-derived neoantigens demonstrate substantially higher immunogenicity than SNV-derived peptides. Mass spectrometry immunopeptidome studies confirm that splicing neoantigens are naturally processed and presented on MHC class I molecules (Li et al., 2024; Kwok et al., 2025; Chai et al., 2022).

Because splicing factor mutations such as SF3B1^{K700E} and SRSF2 hotspot mutations recur across patients and influence splice-site choice in a reproducible manner, some resulting neojunctions are shared among individuals with the same mutation and compatible HLA types. Recent large-scale analyses have identified shared splicing neoantigens in up to 90% of melanoma patients (Li et al., 2024), and tumor-wide neojunctions from genes such as *GNAS* and *RPL22* that are spatially and temporally conserved across entire tumors, including at metastatic sites (Kwok et al., 2025). In myeloid neoplasms, the SF3B1^{K700E} mutation itself generates a neoantigen that has been validated as a CD8⁺ T cell target with nanomolar functional avidity (Biernacki et al., 2025). Such “public” neoantigens raise the possibility of partially off-the-shelf vaccine or T-cell receptor (TCR) therapies that target groups of patients rather than requiring fully individualized designs.

7.6 Computational discovery pipelines

Identifying splicing-derived neoantigens requires integrative pipelines that connect RNA processing to antigen presentation. Multiple specialized computational tools have been developed to systematically predict splicing neoantigens from RNA-seq data. SNAF (Splicing Neo Antigen Finder) combines splice junction calling with deep learning-based immunogenicity prediction (DeepImmuno), tumor specificity ranking (BayesTS), and splicing regulator identification (RNA-SPRINT) (Li et al., 2024). Other validated tools include SpliceMutr for pan-cancer analysis (Palmer et al., 2024), NeoSplice (Chai et al., 2022), and ASNEO for personalized alternative splicing-based neoantigen identification (Zhang et al., 2020). These workflows reconstruct altered open reading frames, enumerate candidate peptides, and filter against normal tissue expression to enrich for tumor-specific sequences.

Downstream modules predict MHC binding, proteasomal processing, and expression levels to prioritize candidates for experimental validation. However, computational predictions remain imperfect: mass spectrometry validation studies

demonstrate that only a subset of predicted neoantigens are actually presented on HLA molecules, highlighting the need for orthogonal experimental confirmation including immunopeptidome profiling and T cell reactivity assays (Li et al., 2024; Chai et al., 2022; Huber et al., 2025; Tretter et al., 2023). Despite these limitations, scalable computational prioritization provides an essential first step to focus resources on the most promising splicing-derived candidates.

7.7 Structural immunoinformatics: predicting recognition

For a candidate peptide to be a useful target, it must be stably presented by MHC molecules and recognized by T cells. Sequence-based predictors such as NetMHCpan estimate peptide-HLA binding affinities by integrating both binding affinity and mass spectrometry-eluted ligand data (Reynisson et al., 2020). However, high predicted affinity alone does not guarantee immunogenicity. Recent advances in structural modeling of peptide-MHC (pMHC) complexes using fine-tuned AlphaFold models achieve sub-angstrom accuracy and can provide additional insight into peptide conformation, anchor residue placement, and the structural distinctiveness of neoantigen-pMHC surfaces from self-peptides (Motmaen et al., 2023; Mikhaylov and Levine, 2023; Glukhov et al., 2024). Crucially, static binding affinity does not always correlate with immunogenicity, and peptide-MHC complex stability can often be a better predictor of T-cell responses (Rasmussen et al., 2016). Molecular dynamics simulations use metrics such as peptide backbone root mean square fluctuation (RMSF, a measure of atomic positional fluctuation over time) to quantify dynamic stability within the MHC groove, which have been linked to differences in TCR recognition and functional responses (Ma et al., 2025; Tomasiak et al., 2022; Bell et al., 2021). In case studies of cancer neoantigens, single amino-acid substitutions, including at anchor positions, can reconfigure peptide-MHC allostery so that the neoantigen populates groove conformations that permissively gate TCR binding, even when MHC binding affinity is similar to the corresponding wild-type peptide (Custodio et al., 2023; Ma et al., 2025). Large-scale analysis of validated neopeptopes has identified the presence of hydrophobic and aromatic residues in the peptide binding core as among the most important features predicting T cell immunogenicity (Borch et al., 2024). In SF3B1-mutant uveal melanoma, splicing-driven shared neoantigens have been shown to elicit robust memory CD8⁺ T-cell responses (Bigot et al., 2021). More recently, tumor-wide public neoantigens derived from recurrent splicing aberrations in genes such as *GNAS* and *RPL22* have been validated to trigger neoantigen-specific CD8⁺ T-cell responses with nanomolar functional avidity and mediate antigen-dependent tumor cell killing (Kwok et al., 2025). These examples illustrate a path toward immunotherapies that exploit the distinctive neoantigen landscape created by splicing factor mutations.

The integration of these AI/MD-driven structural insights allows for a transition from qualitative neoantigen identification to quantitative clinical prediction. By moving beyond simple sequence enumeration to assess the structural “foreignness” and MHC-binding stability of the splicing-derived repertoire, we can derive a high-confidence “integrated neoantigen burden” score. This

score serves as a critical prognostic biomarker in precision oncology, enabling the stratification of patients based on their likelihood of responding to ICB and providing a molecular rationale for the observed clinical variability in immunotherapy outcomes.

8 Conclusion

The spliceosome is increasingly shifting from a perceived “undruggable” black box of prohibitive complexity to a system with tractable, dynamic vulnerabilities, that serve as the foundation for novel diagnostic and prognostic frameworks. As discussed in this review, the combination of high-resolution structural biology, physics-based simulations, and emerging generative AI methods is beginning to narrow the gap between static atomic snapshots and the transient, fluid reality of RNA processing.

Somatic mutations in *SF3B1*, *U2AF1*, and *SRSF2* are now understood not simply as loss-of-function events, but as specific structural perturbations that rewire splicing decisions, conceptually reshaping the underlying energy landscapes to favor aberrant recognition events. Exploiting these defects will likely require the coordinated use of complementary computational approaches. Generative foundation models such as AlphaFold3 and Boltz-2 can, in favorable cases, rapidly propose structural hypotheses for complex ribonucleoprotein interfaces, while atomistic simulations and enhanced sampling methods remain essential for resolving “breathing” motions and cryptic pockets that are inaccessible to static models.

As generative AI continues to evolve, its specific architectures and tools will undoubtedly change. However, the broader shift towards using these models and simulations to analyse dynamic conformational ensembles rather than isolated structures is set to become a key theme in spliceosome-focused drug discovery.

Looking ahead, increasing structural and dynamic insight into mutant spliceosomes has the potential to support a range of therapeutic strategies. The mapping of mutation-specific conformational signatures has the potential to facilitate the design of context-sensitive small molecules that preferentially engage transient allosteric sites, or “molecular glues”, which could stabilise otherwise weak regulatory interactions within the splicing machinery. At the same time, the recurrent and partially predictable nature of some splicing alterations may reveal a pool of candidate tumour-specific neoantigens. As discussed, these may serve as high-value prognostic biomarkers, where the calculated ‘neoantigenic burden’, refined by AI-driven structural validation, could predict patient response to immunotherapy in a clinical setting.

Although substantial experimental, computational, and clinical challenges remain, the convergence of dynamic structural determination with AI-enabled design offers a realistic path towards multi-omics integration for precision oncology. This synergy has the potential to translate an increasingly detailed

understanding of RNA processing into clinically actionable diagnostic and prognostic biomarkers and therapeutic strategies.

Author contributions

JS: Writing – original draft, Writing – review and editing. AK: Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the Swiss National Science Foundation (SNF) (Grant number 222878).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. During the preparation of this work, the authors used Perplexity AI Pro to assist with literature searches, source identification, and summarizing background research findings. Additionally, Google Gemini 3 Pro was used to improve language clarity and readability of draft text, and to enhance the visual design of schematic figures based on original concepts and layouts designed by the authors. All AI-generated content was thoroughly reviewed, fact-checked, and edited by the authors to ensure accuracy, scientific validity, and originality. The authors take full responsibility for the content of this publication.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. doi:10.1038/s41586-024-07487-w

Amorello, A. N., Chandrashekar Reddy, G., Melillo, B., Cravatt, B. F., Ghosh, A. K., and Jurica, M. S. (2025). SF3B1 thermostability as an assay for splicing inhibitor interactions. *J. Biol. Chem.* 301, 108135. doi:10.1016/j.jbc.2024.108135

- Anczukow, O., Allain, F. H.-T., Angarola, B. L., Black, D. L., Brooks, A. N., Cheng, C., et al. (2024). Steering research on mRNA splicing in cancer towards clinical translation. *Nat. Rev. Cancer* 24, 887–905. doi:10.1038/s41568-024-00750-2
- Aranganathan, A., Gu, X., Wang, D., Vani, B. P., and Tiwary, P. (2025). Modeling Boltzmann-weighted structural ensembles of proteins using artificial intelligence-based methods. *Curr. Opin. Struct. Biol.* 91, 103000. doi:10.1016/j.sbi.2025.103000
- Aristoff, D., Copperman, J., Simpson, G., Webber, R. J., and Zuckerman, D. M. (2023). Weighted ensemble: recent mathematical developments. *J. Chem. Phys.* 158, 014108. doi:10.1063/5.0110873
- Badar, T., Vanegas, Y. A. M., Nanaa, A., Foran, J. M., Al-Kali, A., Mangaonkar, A., et al. (2023). U2AF1 pathogenic variants in myeloid neoplasms and precursor states: distribution of co-mutations and prognostic heterogeneity. *Blood Cancer J.* 13, 149. doi:10.1038/s41408-023-00922-7
- Bak-Gordon, P., and Manley, J. L. (2025). SF3B1: from core splicing factor to oncogenic driver. *RNA* 31, 314–332. doi:10.12611/rna.080368.124
- Baltrukovich, H., and Bartos, P. (2023). RNA-protein complexes and force field polarizability. *Front. Chem.* 11, 1217506. doi:10.3389/fchem.2023.1217506
- Bell, D. R., Domeniconi, G., Yang, C.-C., Zhou, R., Zhang, L., and Cong, G. (2021). Dynamics-based peptide-MHC binding optimization by a convolutional variational autoencoder: a use-case model for CASTELO. *J. Chem. Theory Comput.* 17, 7962–7971. doi:10.1021/acs.jctc.1c00870
- Bemelmans, M. P., Cournia, Z., Damm-Ganamet, K. L., Gervasio, F. L., and Pande, V. (2025). Computational advances in discovering cryptic pockets for drug discovery. *Curr. Opin. Struct. Biol.* 90, 102975. doi:10.1016/j.sbi.2024.102975
- Bewersdorf, J. P., Stahl, M., Taylor, J., Mi, X., Chandhok, N. S., Watts, J., et al. (2023). E7820, an anti-cancer sulfonamide, degrades RBM39 in patients with splicing factor mutant myeloid malignancies: a phase II clinical trial. *Leukemia* 37, 2512–2516. doi:10.1038/s41375-023-02050-4
- Biernacki, M. A., Lok, J., Foster, K. A., Cummings, C., Busch, S., Black, R. G., et al. (2025). SF3B1K700E neoantigen is a CD8+ T-cell target shared across human myeloid neoplasms. *Cancer Immunol. Res.* 13, 1391–1404. doi:10.1158/2326-6066.CIR-24-0091
- Bigot, J., Lalanne, A. L., Lucibello, F., Gueguen, P., Houy, A., Dayot, S., et al. (2021). Splicing patterns in SF3B1-Mutated uveal melanoma generate shared immunogenic tumor-specific neopeptides. *Cancer Discov.* 11, 1938–1951. doi:10.1158/2159-8290.CD-20-0555
- Bland, P., Saville, H., Wai, P. T., Curnow, L., Muirhead, G., Nieminiusz, J., et al. (2023). SF3B1 hotspot mutations confer sensitivity to PARP inhibition by eliciting a defective replication stress response. *Nat. Genet.* 55, 1311–1323. doi:10.1038/s41588-023-01460-5
- Borch, A., Carri, I., Reynisson, B., Alvarez, H. M. G., Munk, K. K., Montemurro, A., et al. (2024). IMPROVE: a feature model to predict neopeptide immunogenicity through broad-scale validation of T-cell recognition. *Front. Immunol.* 15, 1360281. doi:10.3389/fimmu.2024.1360281
- Borišek, J., Saltalamacchia, A., Spinello, A., and Magistrato, A. (2020). Exploiting Cryo-EM structural information and all-atom simulations to decrypt the molecular mechanism of splicing modulators. *J. Chem. Inf. Model.* 60, 2510–2521. doi:10.1021/acs.jcim.9b00635
- Borišek, J., Casalino, L., Saltalamacchia, A., Mays, S. G., Malcovati, L., and Magistrato, A. (2021). Atomic-level mechanism of Pre-mRNA splicing in health and disease. *Accounts Chem. Res.* 54, 144–154. doi:10.1021/acs.accounts.0c00578
- Borišek, J., Aupič, J., and Magistrato, A. (2023). Establishing the catalytic and regulatory mechanism of β -based machineries. *WIREs Comput. Mol. Sci.* 13, e1643. doi:10.1002/wcms.1643
- Bose, S., Kilinc, C., and Dickson, A. (2025). Markov state models with weighted ensemble simulation: how to eliminate the trajectory merging bias. *J. Chem. Theory Comput.* 21, 1805–1816. doi:10.1021/acs.jctc.4c01141
- Buttenschoen, M., Morris, G. M., and Deane, C. M. (2024). PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* 15, 3130–3139. doi:10.1039/D3SC04185A
- Chai, S., Smith, C. C., Kochar, T. K., Hunsucker, S. A., Beck, W., Olsen, K. S., et al. (2022). NeoSplice: a bioinformatics method for prediction of splice variant neoantigens. *Bioinforma. Adv.* 2, vbac032. doi:10.1093/bioadv/vbac032
- Chai, D., Boitreaud, J., Dent, J., McPartlon, M., Meier, J., Reis, V., et al. (2024). Chai-1: decoding the molecular interactions of life. *bioRxiv*. 10.1101/2024.10.10.615955.
- Chan, W. K., Carlson, H. A., and Traynor, J. R. (2023). Application of mixed-solvent molecular dynamics simulations for prediction of allosteric sites on G protein-coupled receptors. *Mol. Pharmacol.* 103, 274–285. doi:10.1124/molpharm.122.000612
- Chen, M., and Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754. doi:10.1038/nrm2777
- Chen, W., He, H., Wang, J., Wang, J., and Chang, C.-e. A. (2023). Uncovering water effects in protein-ligand recognition: importance in the second hydration shell and binding kinetics. *Phys. Chem. Chem. Phys.* 25, 2098–2109. doi:10.1039/D2CP04584B
- Chong, L. T., and Zuckerman, D. M. (2025). Weighted ensemble simulation: advances in methods, software, and applications. *WIREs Comput. Mol. Sci.* 15, e70055. doi:10.1002/wcms.70055
- Chyzy, P., Kulik, M., Shinobu, A., Re, S., Sugita, Y., and Trylska, J. (2024). Molecular dynamics in multidimensional space explains how mutations affect the association path of neomycin to a riboswitch. *Proc. Natl. Acad. Sci.* 121, e2317197121. doi:10.1073/pnas.2317197121
- Corrionero, A., Miñana, B., and Valcárcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes & Dev.* 25, 445–459. doi:10.1101/gad.2014311
- Cretu, C., Schmitzová, J., Ponce-Salvatierra, A., Dybkov, O., De Laurentiis, E., Sharma, K., et al. (2016). Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* 64, 307–319. doi:10.1016/j.molcel.2016.08.036
- Cui, X., Ge, L., Chen, X., Lv, Z., Wang, S., Zhou, X., et al. (2025). Beyond static structures: protein dynamic conformations modeling in the post-AlphaFold era. *Briefings Bioinforma.* 26, bbaf340. doi:10.1093/bib/bbaf340
- Custodio, J. M., Ayres, C. M., Rosales, T. J., Brambley, C. A., Arbuiso, A. G., Landau, L. M., et al. (2023). Structural and physical features that distinguish tumor-controlling from inactive cancer neopeptides. *Proc. Natl. Acad. Sci.* 120, e2312057120. doi:10.1073/pnas.2312057120
- Daubner, G. M., Cléry, A., Jayne, S., Stevenin, J., and Allain, F. H. (2012). A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.* 31, 162–174. doi:10.1038/emboj.2011.367
- Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 378, 49–56. doi:10.1126/science.add2187
- Dauparas, J., Lee, G. R., Pecoraro, R., An, L., Anishchenko, I., Glasscock, C., et al. (2025). Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* 22, 717–723. doi:10.1038/s41592-025-02626-1
- Desai, D., Kantliwala, S. V., Vybhavi, J., Ravi, R., Patel, H., and Patel, J. (2024). Review of AlphaFold 3: transformative advances in drug design and therapeutics. *Cureus* 16, e63646. doi:10.7759/cureus.63646
- Dolatshad, H., Pellagatti, A., Liberante, F. G., Llorian, M., Repapi, E., Steeples, V., et al. (2022). Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. *Leukemia* 30, 2322–2331. doi:10.1038/leu.2016.149
- Dörner, K., and Hondele, M. (2024). The story of RNA unfolded: the molecular function of DEAD- and DEXH-Box ATPases and their complex relationship with membraneless organelles. *Annu. Rev. Biochem.* 93, 79–108. doi:10.1146/annurev-biochem-052521-121259
- Dvigne, H., Kim, E., Abdel-Wahab, O., and Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* 16, 413–430. doi:10.1038/nrc.2016.51
- Eldfors, S., Rai, S., Sharma, V., Hossan, T., Cabrera, C., Bertino, A., et al. (2024). 3067 – ATR/CHK1/WEE1 dependency in SRSF2-mutated MDS/AML. *Exp. Hematol.* 137, 104389. doi:10.1016/j.exphem.2024.104389
- Ellis, M. J., Ding, L., Shen, D., Luo, J., Suman, V. J., Wallis, J. W., et al. (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486, 353–360. doi:10.1038/nature11143
- Eymin, B. (2021). Targeting the spliceosome machinery: a new therapeutic axis in cancer? *Biochem. Pharmacol.* 189, 114039. doi:10.1016/j.bcp.2020.114039
- Fang, Z., Ran, H., Zhang, Y., Chen, C., Lin, P., Zhang, X., et al. (2025). AlphaFold 3: an unprecedented opportunity for fundamental research and drug development. *Precis. Clin. Med.* 8, pbaf015. doi:10.1093/pmedi/pbaf015
- Favor, A., Quijano, R., Chernova, E., Kubaney, A., Weidle, C., Esler, M. A., et al. (2025). De novo design of RNA and nucleoprotein complexes. *bioRxiv*. 10.1101/2025.10.01.679929.
- Fica, S. M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. & Mol. Biol.* 24, 791–799. doi:10.1038/nsmb.3463
- Fica, S. M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., et al. (2013). RNA catalyzes nuclear pre-mRNA splicing. *Nature* 503, 229–234. doi:10.1038/nature12734
- Fong, J. Y., Pignata, L., Goy, P.-A., Kawabata, K. C., Lee, S. C.-W., Koh, C. M., et al. (2019). Therapeutic targeting of RNA splicing catalysis through inhibition of protein arginine methylation. *Cancer Cell* 36, 194–209.e9. doi:10.1016/j.ccell.2019.07.003
- Gapsys, V., Kopec, W., Matthes, D., and De Groot, B. L. (2024). Biomolecular simulations at the exascale: from drug design to organelles and beyond. *Curr. Opin. Struct. Biol.* 88, 102887. doi:10.1016/j.sbi.2024.102887
- García-Manero, G., Platzbecker, U., Lim, K.-H., Nowakowski, G., Abdel-Wahab, O., Kantarjian, H., et al. (2024). Research and clinical updates on IRAK4 and its roles in inflammation and malignancy: themes and highlights from the 1st symposium on IRAK4 in cancer. *Front. Hematol.* 3, 1339870. doi:10.3389/frhem.2024.1339870
- Glukhov, E., Kalitin, D., Stepanenko, D., Zhu, Y., Nguyen, T., Jones, G., et al. (2024). MHC-Fine: fine-tuned AlphaFold for precise MHC-peptide complex prediction. *Biophysical J.* 123, 2902–2909. doi:10.1016/j.bpj.2024.05.011
- Grotz, K. K., Cruz-León, S., and Schwierz, N. (2021). Optimized magnesium force field parameters for biomolecular simulations with accurate solvation, ion-binding, and

- water-exchange properties. *J. Chem. Theory Comput.* 17, 2530–2540. doi:10.1021/acs.jctc.0c01281
- Han, T., Goralski, M., Gaskill, N., Capota, E., Kim, J., Ting, T. C., et al. (2017). Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science* 356, eaal3755. doi:10.1126/science.aal3755
- Harbour, J. W., Onken, M. D., Roberson, E. D. O., Duan, S., Cao, L., Worley, L. A., et al. (2010). Frequent mutation of *BAP1* in metastasizing uveal melanomas. *Science* 330, 1410–1413. doi:10.1126/science.1194472
- Harbour, J. W., Roberson, E. D. O., Anbunathan, H., Onken, M. D., Worley, L. A., and Bowcock, A. M. (2013). Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat. Genet.* 45, 133–135. doi:10.1038/ng.2523
- Havens, M. A., and Hastings, M. L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res.* 44, 6549–6563. doi:10.1093/nar/gkw533
- Helleman, E., and Durrant, J. D. (2023). Worth the weight: Sub-Pocket Explorer (SubPEX), a weighted ensemble method to enhance binding-pocket conformational sampling. *J. Chem. Theory Comput.* 19, 5677–5689. doi:10.1021/acs.jctc.3c00478
- Hoff, F. W., Blum, W. G., Huang, Y., Welkie, R. L., Swords, R. T., Traer, E., et al. (2024). Beat-AML 2024 ELN-refined risk stratification for older adults with newly diagnosed AML given lower-intensity therapy. *Blood Adv.* 8, 5297–5305. doi:10.1182/bloodadvances.2024013685
- Hollingsworth, S. A., and Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron* 99, 1129–1143. doi:10.1016/j.neuron.2018.08.011
- Hong, D. S., Kurzrock, R., Naing, A., Wheler, J. J., Falchook, G. S., Schiffman, J. S., et al. (2014). A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) splicing inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors. *Investig. New Drugs* 32, 436–444. doi:10.1007/s10637-013-0046-5
- Hooten, M., Patel, H., Shao, Y., Singh, R. K., and Dutt, M. (2025). Applications of enhanced sampling methods to biomolecular self-assembly: a review. *J. Phys. Condens. Matter* 37, 303001. doi:10.1088/1361-648X/ade650
- Huang, H. H., Ferguson, I. D., Thornton, A. M., Bastola, P., Lam, C., Lin, Y.-H. T., et al. (2020). Proteasome inhibitor-induced modulation reveals the spliceosome as a specific therapeutic vulnerability in multiple myeloma. *Nat. Commun.* 11, 1931. doi:10.1038/s41467-020-15521-4
- Huber, F., Arnaud, M., Stevenson, B. J., Michaux, J., Benedetti, F., Thevenet, J., et al. (2025). A comprehensive proteogenomic pipeline for neoantigen discovery to advance personalized cancer immunotherapy. *Nat. Biotechnol.* 43, 1360–1372. doi:10.1038/s41587-024-02420-y
- Ilgan, J. O., Ramakrishnan, A., Hayes, B., Murphy, M. E., Zebari, A. S., Bradley, P., et al. (2015). *U2AF1* mutations alter splice site recognition in hematological malignancies. *Genome Res.* 25, 14–26. doi:10.1101/gr.181016.114
- Inoue, D., Chew, G.-L., Liu, B., Michel, B. C., Pangallo, J., D'Avino, A. R., et al. (2019). Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* 574, 432–436. doi:10.1038/s41586-019-1646-9
- Jiang, W., and Chen, L. (2021). Alternative splicing: human disease and quantitative analysis from high-throughput sequencing. *Comput. Struct. Biotechnol. J.* 19, 183–195. doi:10.1016/j.csbj.2020.12.009
- Jiang, M., Chen, M., Liu, Q., Jin, Z., Yang, X., and Zhang, W. (2023). SF3B1 mutations in myelodysplastic syndromes: a potential therapeutic target for modulating the entire disease process. *Front. Oncol.* 13, 1116438. doi:10.3389/fonc.2023.1116438
- Jin, P., Wang, X., Jin, Q., Zhang, Y., Shen, J., Jiang, G., et al. (2024). Mutant U2AF1-Induced mis-splicing of mrna translation genes confers resistance to chemotherapy in acute myeloid leukemia. *Cancer Res.* 84, 1583–1596. doi:10.1158/0008-5472.CAN-23-2543
- Jing, B., Berger, B., and Jaakkola, T. (2024). AlphaFold meets flow matching for generating protein ensembles. arXiv. 10.48550/ARXIV.2402.04845.
- Joshi, C. K., Jamsab, A. R., Viñas, R., Harris, C., Mathis, S. V., Morehead, A., et al. (2024). gRNAde: Geometric deep learning for 3D RNA inverse design. bioRxiv. 10.1101/2024.03.31.587283.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kang, S., Lee, D., Hwang, G., Lee, K., and Kang, M. (2025). ASOptimizer: optimizing chemical diversity of antisense oligonucleotides through deep learning. *Nucleic Acids Res.* 53, W39–W44. doi:10.1093/nar/gkaf392
- Kato, M., Han, T., Xie, S., Shi, K., Du, X., Wu, L., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149, 753–767. doi:10.1016/j.cell.2012.04.017
- Kim, E., Ilagan, J., Liang, Y., Daubner, G., Lee, S.-W., Ramakrishnan, A., et al. (2015). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* 27, 617–630. doi:10.1016/j.ccr.2015.04.006
- Kim, S. P., Srivatsan, S. N., Chavez, M., Shirai, C. L., White, B. S., Ahmed, T., et al. (2021). Mutant U2AF1-induced alternative splicing of H2afy (macroH2A1) regulates B-lymphopoiesis in mice. *Cell Rep.* 36, 109626. doi:10.1016/j.celrep.2021.109626
- Kohsaka, S., Yagishita, S., Shirai, Y., Matsuno, Y., Ueno, T., Kojima, S., et al. (2024). A molecular glue RBM39-degrader induces synthetic lethality in cancer cells with homologous recombination repair deficiency. *Npj Precis. Oncol.* 8, 117. doi:10.1038/s41698-024-00610-0
- Kononov, K. A., Unarta, I. C., Cao, S., Goonetilleke, E. C., and Huang, X. (2021). Markov state models to study the functional dynamics of proteins in the wake of machine learning. *JACS Au* 1, 1330–1341. doi:10.1021/jacsau.1c00254
- Koseki, J., Motono, C., Yanagisawa, K., Kudo, G., Yoshino, R., Hirokawa, T., et al. (2025). CrypToth: cryptic pocket detection through mixed-solvent molecular dynamics simulations-based topological data analysis. *J. Chem. Inf. Model.* 65, 5567–5575. doi:10.1021/acs.jcim.4c02111
- Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., et al. (2024). Generalized biomolecular modeling and design with RoseTTAFold all-atom. *Science* 384, eadl2528. doi:10.1126/science.adl2528
- Krokidis, M. G., Koumadorakis, D. E., Lazaros, K., Ivantsik, O., Exarchos, T. P., Vrahasis, A. G., et al. (2025). AlphaFold3: an overview of applications and performance insights. *Int. J. Mol. Sci.* 26, 3671. doi:10.3390/ijms26083671
- Kuzmanic, A., Bowman, G. R., Juarez-Jimenez, J., Michel, J., and Gervasio, F. L. (2020). Investigating cryptic binding sites by molecular dynamics simulations. *Accounts Chem. Res.* 53, 654–661. doi:10.1021/acs.accounts.9b00613
- Kwok, D. W., Stevers, N. O., Etxeberria, I., Nejo, T., Colton Cove, M., Chen, L. H., et al. (2025). Tumour-wide RNA splicing aberrations generate actionable public neoantigens. *Nature* 639, 463–473. doi:10.1038/s41586-024-08552-0
- Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci.* 99, 12562–12566. doi:10.1073/pnas.202427399
- Lewis, S., Hempel, T., Jiménez-Luna, J., Gastegger, M., Xie, Y., Foong, A. Y. K., et al. (2025). Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* 389, eadv9817. doi:10.1126/science.adv9817
- Li, B., Zou, D., Yang, S., Ouyang, G., and Mu, Q. (2020). Prognostic significance of *U2AF1* mutations in myelodysplastic syndromes: a meta-analysis. *J. Int. Med. Res.* 48, 0300060519891013. doi:10.1177/0300060519891013
- Li, G., Mahajan, S., Ma, S., Jeffery, E. D., Zhang, X., Bhattacharjee, A., et al. (2024). Splicing neoantigen discovery with SNAF reveals shared targets for cancer immunotherapy. *Sci. Transl. Med.* 16, eade2886. doi:10.1126/scitranslmed.ade2886
- Liang, Y., Tebaldi, T., Rejeski, K., Joshi, P., Stefani, G., Taylor, A., et al. (2018). SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. *Leukemia* 32, 2659–2671. doi:10.1038/s41375-018-0152-7
- Lin, S., Hong, L., Wei, D.-Q., and Xiong, Y. (2024). Deep learning facilitates efficient optimization of antisense oligonucleotide drugs. *Mol. Ther. - Nucleic Acids* 35, 102208. doi:10.1016/j.omtn.2024.102208
- Liu, X., Devadiga, S. A., Stanley, R. F., Morrow, R. M., Janssen, K. A., Quesnel-Vallières, M., et al. (2024). A mitochondrial surveillance mechanism activated by SRSF2 mutations in hematologic malignancies. *J. Clin. Investigation* 134, e175619. doi:10.1172/JCI1175619
- Liu, B., Boysen, J. G., Unarta, I. C., Du, X., Li, Y., and Huang, X. (2025). Exploring transition states of protein conformational changes via out-of-distribution detection in the hyperspherical latent space. *Nat. Commun.* 16, 349. doi:10.1038/s41467-024-55228-4
- Lu, Z., Jiang, P., and Xing, Y. (2012). Genetic variation of pre-mRNA alternative splicing in human populations. *WIREs RNA* 3, 581–592. doi:10.1002/wrna.120
- Lv, X., Sun, X., Gao, Y., Song, X., Hu, X., Gong, L., et al. (2025). Targeting RNA splicing modulation: new perspectives for anticancer strategy? *J. Exp. & Clin. Cancer Res.* 44, 32. doi:10.1186/s13046-025-03279-w
- Ma, J., Ayres, C. M., Brambley, C. A., Chandran, S. S., Rosales, T. J., Perera, W. W. J. G., et al. (2025). Dynamic allostery in the peptide/MHC complex enables TCR neoantigen selectivity. *Nat. Commun.* 16, 849. doi:10.1038/s41467-025-56004-8
- Mackereth, C. D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., et al. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* 475, 408–411. doi:10.1038/nature10171
- Malcovati, L., Stevenson, K., Papaemmanuil, E., Neuberger, D., Bejar, R., Boulton, J., et al. (2020). *SF3B1*-mutant MDS as a distinct disease subtype: a proposal from the international working group for the prognosis of MDS. *Blood* 136, 157–170. doi:10.1182/blood.2020004850
- Malhotra, Y., John, J., Yadav, D., Sharma, D., Vanshika, Rawal, K., et al. (2025). Advancements in protein structure prediction: a comparative overview of AlphaFold and its derivatives. *Comput. Biol. Med.* 188, 109842. doi:10.1016/j.combiomed.2025.109842
- Mehdi, S., Smith, Z., Herron, L., Zou, Z., and Tiwary, P. (2024). Enhanced sampling with machine learning. *Annu. Rev. Phys. Chem.* 75, 347–370. doi:10.1146/annurev-physchem-083122-125941
- Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., et al. (2023). UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* 32, e4792. doi:10.1002/pro.4792
- Migens, P. F., Serrano-Chacón, I., Orozco, M., and Battistini, F. (2025). Success and limitations of current force fields for the description of RNA–ligand complexes. *J. Phys. Chem. B* 129, 11679–11692. doi:10.1021/acs.jpcc.5c05262

- Mikhailov, V., and Levine, A. J. (2023). Accurate modeling of peptide-MHC structures with AlphaFold. *bioRxiv*. 10.1101/2023.03.06.531396.
- Mlýnský, V., Kůhrová, P., Pykal, M., Krepl, M., Stadlbauer, P., Otyepka, M., et al. (2025). Can we ever develop an ideal RNA force field? Lessons learned from simulations of the UUCG RNA tetraloop and other systems. *J. Chem. Theory Comput.* 21, 4183–4202. doi:10.1021/acs.jctc.4c01357
- Motmaen, A., Dauparas, J., Baek, M., Abedi, M. H., Baker, D., and Bradley, P. (2023). Peptide-binding specificity prediction using fine-tuned protein structure prediction networks. *Proc. Natl. Acad. Sci.* 120, e2216697120. doi:10.1073/pnas.2216697120
- Mukhopadhyay, S., and Chakrabarty, S. (2025). Mixed-Solvent molecular dynamics simulation reveals a druggable allosteric pocket in the PCSK9 C-Terminal domain. *J. Phys. Chem. B* 129, 11737–11744. doi:10.1021/acs.jpcc.5c05530
- Mukhopadhyay, S., Sinha, K., and Chakrabarty, S. (2025). Harnessing allostery to modulate protein-protein interactions: from function to therapeutic innovations. *J. Mol. Biol.* 437, 169382. doi:10.1016/j.jmb.2025.169382
- Muscat, S., Martino, G., Manigrasso, J., Marcia, M., and De Vivo, M. (2024). On the power and challenges of atomistic molecular dynamics to investigate RNA molecules. *J. Chem. Theory Comput.* doi:10.1021/acs.jctc.4c00773
- Nguyen, H. D., Leong, W. Y., Li, W., Reddy, P. N., Sullivan, J. D., Walter, M. J., et al. (2018). Spliceosome mutations induce R-loop-associated sensitivity to ATR inhibition in myelodysplastic syndromes. *Cancer Res.* 78, 5363–5374. doi:10.1158/0008-5472.CAN-17-3970
- Nori, D., Parsan, A., Uhler, C., and Jin, W. (2025). BindEnergyCraft: casting protein structure predictors as energy-based models for binder design. *arXiv*. doi:10.48550/ARXIV.2505.21241
- Ohr, T., Prior, M., Dannenberg, J., Odenwälder, P., Dybkov, O., Rasche, N., et al. (2012). Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* 18, 1244–1256. doi:10.1261/rna.033316.112
- Okeyo-Owuor, T., White, B. S., Chatrikhi, R., Mohan, D. R., Kim, S., Griffith, M., et al. (2015). U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia* 29, 909–917. doi:10.1038/leu.2014.303
- Pacesa, M., Nickel, L., Schellhaas, C., Schmidt, J., Pyatova, E., Kissling, L., et al. (2025). One-shot design of functional protein binders with BindCraft. *Nature* 646, 483–492. doi:10.1038/s41586-025-09429-6
- Palmer, T., Kessler, M. D., Shao, X. M., Balan, A., Yarchoan, M., Zaidi, N., et al. (2024). SpliceMutr enables pan-cancer analysis of splicing-derived neoantigen burden in tumors. *Cancer Res. Commun.* 4, 3137–3150. doi:10.1158/2767-9764.CRC-23-0309
- Papaemmanuil, E., Cazzola, M., Boulton, J., Malcovati, L., Vyas, P., Bowen, D., et al. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* 365, 1384–1395. doi:10.1056/NEJMoa1103283
- Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Thaler, S., Somnath, V. R., et al. (2025). Boltz-2: towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025.06.14.659707. doi:10.1101/2025.06.14.659707
- Peng, S., Long, M., Chen, Q., Yin, Z., Zeng, C., Zhang, W., et al. (2025). Perspectives on cancer therapy—synthetic lethal precision medicine strategies, molecular mechanisms, therapeutic targets and current technical challenges. *Cell Death Discov.* 11, 179. doi:10.1038/s41420-025-02418-8
- Penić, R. J., Vlašić, T., Huber, R. G., Wan, Y., and Šikić, M. (2025). RiNALMo: general-purpose RNA language models can generalize well on structure prediction tasks. *Nat. Commun.* 16, 5671. doi:10.1038/s41467-025-60872-5
- Pokorná, P., Aupič, J., Fica, S. M., and Magistrato, A. (2025). Decoding spliceosome dynamics through computation and experiment. *Chem. Rev.* 125, 9807–9833. doi:10.1021/acs.chemrev.5c00374
- Qiu, J., Zhou, B., Thol, F., Zhou, Y., Chen, L., Shao, C., et al. (2016). Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators. *RNA* 22, 1535–1549. doi:10.1261/rna.056101.116
- Quesada, V., Conde, L., Villamor, N., Ordóñez, G. R., Jares, P., Bassaganyas, L., et al. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* 44, 47–52. doi:10.1038/ng.1032
- Raghunathan, P. L., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* 8, 847–855. doi:10.1016/S0960-9822(07)00345-4
- Rahman, M. A., Lin, K.-T., Bradley, R. K., Abdel-Wahab, O., and Krainer, A. R. (2020). Recurrent SRSF2 mutations in MDS affect both splicing and NMD. *Genes & Dev.* 34, 413–427. doi:10.1101/gad.32270.119
- Rakesh, R., Joseph, A. P., Bhaskara, R. M., and Srinivasan, N. (2016). Structural and mechanistic insights into human splicing factor SF3b complex derived using an integrated approach guided by the cryo-EM density maps. *RNA Biol.* 13, 1025–1040. doi:10.1080/15476286.2016.1218590
- Rasmussen, M., Fenoy, E., Harndahl, M., Kristensen, A. B., Nielsen, I. K., Nielsen, M., et al. (2016). Pan-specific prediction of Peptide-MHC class I complex stability, a correlate of T cell immunogenicity. *J. Immunol.* 197, 1517–1524. doi:10.4049/jimmunol.1600582
- Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 48, W449–W454. doi:10.1093/nar/gkaa379
- Ross, G. A., Lu, C., Scarabelli, G., Albanese, S. K., Houang, E., Abel, R., et al. (2023). The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. *Commun. Chem.* 6, 222. doi:10.1038/s42004-023-01019-9
- Rozza, R., Janoš, P., and Magistrato, A. (2023). Assessing the binding mode of a splicing modulator stimulating Pre-mRNA binding to the plastic U2AF2 splicing factor. *J. Chem. Inf. Model.* 63, 7508–7517. doi:10.1021/acs.jcim.3c01204
- Sala, D., Engelberger, F., Mchaurab, H., and Meiler, J. (2023). Modeling conformational states of proteins with AlphaFold. *Curr. Opin. Struct. Biol.* 81, 102645. doi:10.1016/j.sbi.2023.102645
- Sampson, J. M., Cannon, D. A., Duan, J., Epstein, J. C., Sergeeva, A. P., Katsamba, P. S., et al. (2024). Robust prediction of relative binding energies for protein-protein complex mutations using free energy perturbation calculations. *J. Mol. Biol.* 436, 168640. doi:10.1016/j.jmb.2024.168640
- Schapiro, M., Halabedian, L., Arrowsmith, C. H., and Harding, R. J. (2024). Big data and benchmarking initiatives to bridge the gap from AlphaFold to drug design. *Nat. Chem. Biol.* 20, 937–940. doi:10.1038/s41589-024-01570-z
- Seiler, M., Yoshimi, A., Darman, R., Chan, B., Keaney, G., Thomas, M., et al. (2018). H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat. Med.* 24, 497–504. doi:10.1038/nm.4493
- Shaw, D. E., Dror, R. O., Salmon, J. K., Grossman, J. P., Mackenzie, K. M., Bank, J. A., et al. (2009). “Millisecond-scale molecular dynamics simulations on anton,” in Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis (Portland Oregon: ACM), Portland, OR, USA, 14–20 November 2009 (IEEE), 1–11.
- Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R., and Kielkopf, C. L. (2006). Structural basis for polypyrimidine tract recognition by the essential Pre-mRNA splicing factor U2AF65. *Mol. Cell* 23, 49–59. doi:10.1016/j.molcel.2006.05.025
- Škrinjar, P., Eberhardt, J., Tauriello, G., Schwede, T., and Durairaj, J. (2025). Have protein-ligand cofolding methods moved beyond memorisation? *bioRxiv*. 10.1101/2025.02.03.636309.
- Smart, A. C., Margolis, C. A., Pimentel, H., He, M. X., Miao, D., Adeegbe, D., et al. (2018). Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* 36, 1056–1058. doi:10.1038/nbt.4239
- Smith, R. D., and Carlson, H. A. (2021). Identification of cryptic binding sites using MixMD with standard and accelerated molecular dynamics. *J. Chem. Inf. Model.* 61, 1287–1299. doi:10.1021/acs.jcim.0c01002
- Smith, M. A., Choudhary, G. S., Pellagatti, A., Choi, K., Bolanos, L. C., Bhagat, T. D., et al. (2019). U2AF1 mutations induce oncogenic IRAK4 isoforms and activate innate immune pathways in myeloid malignancies. *Nat. Cell Biol.* 21, 640–650. doi:10.1038/s41556-019-0314-5
- Souza, P. C. T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünwald, F., et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* 18, 382–388. doi:10.1038/s41592-021-01098-3
- Spinello, A., Borišek, J., Malcovati, L., and Magistrato, A. (2021). Investigating the molecular mechanism of H3B-8800: a splicing modulator inducing preferential lethality in spliceosome-mutant cancers. *Int. J. Mol. Sci.* 22, 11222. doi:10.3390/ijms222011222
- Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R. A., et al. (2018). RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* 118, 4177–4338. doi:10.1021/acs.chemrev.7b00427
- Stark, H., Faltings, F., Choi, M., Xie, Y., Hur, E., O'Donnell, T., et al. (2025). BoltzGen: toward universal binder design. *bioRxiv*. 10.1101/2025.11.20.689494.
- Steensma, D. P., Wermke, M., Klimek, V. M., Greenberg, P. L., Font, P., Komrokji, R. S., et al. (2019). Results of a clinical trial of H3B-8800, a splicing modulator, in patients with myelodysplastic syndromes (MDS), acute myeloid leukemia (AML) or chronic myelomonocytic leukemia (CMML). *Blood* 134, 673. doi:10.1182/blood-2019-123854
- Su, M., Fleischer, T., Grosheva, I., Horev, M. B., Olszewska, M., Mattioli, C. C., et al. (2024). Targeting SRSF2 mutations in leukemia with RKI-1447: a strategy to impair cellular division and nuclear structure. *iScience* 27, 109443. doi:10.1016/j.isci.2024.109443
- Sun, Q., Wang, H., Xie, J., Wang, L., Mu, J., Li, J., et al. (2025). Computer-aided drug discovery for undruggable targets. *Chem. Rev.* 125, 6309–6365. doi:10.1021/acs.chemrev.4c00969
- Takada, S. (2019). Go model revisited. *Biophysics Physicobiology* 16, 248–255. doi:10.2142/biophysico.16.0_248
- Takehima, Y. (2025). Expansion of splice-switching therapy with antisense oligonucleotides. *Int. J. Mol. Sci.* 26, 2270. doi:10.3390/ijms26052270
- Tholen, J., and Galej, W. (2022). Structural studies of the spliceosome: bridging the gaps. *Curr. Opin. Struct. Biol.* 77, 102461. doi:10.1016/j.sbi.2022.102461
- Tomasiaik, L., Karch, R., and Schreiner, W. (2022). Conformational flexibility of a free and TCR-bound pMHC-I protein investigated by long-term molecular dynamics simulations. *BMC Immunol.* 23, 36. doi:10.1186/s12865-022-00510-7

- Tretter, C., De Andrade Krätzig, N., Pecoraro, M., Lange, S., Seifert, P., Von Frankenberg, C., et al. (2023). Proteogenomic analysis reveals RNA as a source for tumor-agnostic neoantigen identification. *Nat. Commun.* 14, 4632. doi:10.1038/s41467-023-39570-7
- Tseng, C.-C., and Obeng, E. A. (2024). RNA splicing as a therapeutic target in myelodysplastic syndromes. *Seminars Hematol.* 61, 431–441. doi:10.1053/j.seminhematol.2024.10.005
- Vajda, S., Beglov, D., Wakefield, A. E., Egbert, M., and Whitty, A. (2018). Cryptic binding sites on proteins: definition, detection, and druggability. *Curr. Opin. Chem. Biol.* 44, 1–8. doi:10.1016/j.cbpa.2018.05.003
- Vani, B. P., Aranganathan, A., Wang, D., and Tiwary, P. (2023). AlphaFold2-RAVE: from sequence to boltzmann ranking. *J. Chem. Theory Comput.* 19, 4351–4354. doi:10.1021/acs.jctc.3c00290
- Vant, J. W., Sarkar, D., Nguyen, J., Baker, A. T., Vermaas, J. V., and Singharoy, A. (2022). Exploring cryo-electron microscopy with molecular dynamics. *Biochem. Soc. Trans.* 50, 569–581. doi:10.1042/BST20210485
- Verkhivker, G., Alshahrani, M., and Gupta, G. (2023). Exploring conformational landscapes and cryptic binding pockets in distinct functional states of the SARS-CoV-2 omicron BA.1 and BA.2 trimers: mutation-induced modulation of protein dynamics and network-guided prediction of variant-specific allosteric binding sites. *Viruses* 15, 2009. doi:10.3390/v15102009
- Vithani, N., Zhang, S., Thompson, J. P., Patel, L. A., Demidov, A., Xia, J., et al. (2024). Exploration of cryptic pockets using enhanced sampling along normal modes: a case study of KRAS^{G12V}. *J. Chem. Inf. Model.* 64, 8258–8273. doi:10.1021/acs.jcim.4c01435
- Von Bülow, S., Tesei, G., Zaidi, F. K., Mittag, T., and Lindorff-Larsen, K. (2025). Prediction of phase-separation propensities of disordered proteins from sequence. *Proc. Natl. Acad. Sci.* 122, e2417920122. doi:10.1073/pnas.2417920122
- Vorländer, M. K., Pacheco-Fiallos, B., and Plaschka, C. (2022). Structural basis of mRNA maturation: time to put it together. *Curr. Opin. Struct. Biol.* 75, 102431. doi:10.1016/j.sbi.2022.102431
- Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi:10.1016/j.cell.2009.02.009
- Wakefield, A. E., Kozakov, D., and Vajda, S. (2022). Mapping the binding sites of challenging drug targets. *Curr. Opin. Struct. Biol.* 75, 102396. doi:10.1016/j.sbi.2022.102396
- Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* 353, 895–904. doi:10.1126/science.aag2235
- Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., Huang, L.-F., et al. (2015). Mechanism of alternative splicing and its regulation. *Biomed. Rep.* 3, 152–158. doi:10.3892/br.2014.407
- Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., et al. (2023). *De novo* design of protein structure and function with RFdiffusion. *Nature* 620, 1089–1100. doi:10.1038/s41586-023-06415-8
- Wayment-Steele, H. K., Ojoawo, A., Otten, R., Apitz, J. M., Pitsawong, W., Hömberger, M., et al. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* 625, 832–839. doi:10.1038/s41586-023-06832-9
- Wheeler, E. C., Martin, B. J. E., Doyle, W. C., Neaher, S., Conway, C. A., Pitton, C. N., et al. (2024). Splicing modulators impair DNA damage response and induce killing of cohesin-mutant MDS and AML. *Sci. Transl. Med.* 16, eade2774. doi:10.1126/scitranslmed.ade2774
- Wu, D., and Feng, L. (2025). Robust prediction of multiple protein conformations with entropy guidance. bioRxiv. 10.1101/2025.04.26.650728.
- Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402, 832–835. doi:10.1038/45590
- Wu, J., Li, D., and Wang, L. (2024). Overview of PRMT1 modulators: inhibitors and degraders. *Eur. J. Med. Chem.* 279, 116887. doi:10.1016/j.ejmech.2024.116887
- Wu, D., Maus, N., Jha, A., Yang, K., Wales-McGrath, B. D., Jewell, S., et al. (2025). Generative modeling for RNA splicing predictions and design. *eLife*. doi:10.7554/eLife.106043.1
- Xing, P., Bak-Gordon, P., Xie, J., Zhang, J., Liu, Z., and Manley, J. L. (2025). SUGP1 loss drives SF3B1 hotspot mutant missplicing in cancer. *Cell Rep.* 44, 116075. doi:10.1016/j.celrep.2025.116075
- Xu, Y.-Z., and Query, C. C. (2007). Competition between the ATPase Prp5 and branch Region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol. Cell* 28, 838–849. doi:10.1016/j.molcel.2007.09.022
- Xu, S., Feng, Q., Qiao, L., Wu, H., Shen, T., Cheng, Y., et al. (2025). Benchmarking all-atom biomolecular structure prediction with FoldBench. *Nat. Commun.* 17, 442. doi:10.1038/s41467-025-67127-3
- Yang, Y. I., Shao, Q., Zhang, J., Yang, L., and Gao, Y. Q. (2019). Enhanced sampling in molecular dynamics. *J. Chem. Phys.* 151, 070902. doi:10.1063/1.5109531
- Yangeliev, D., and Ozkan, S. B. (2025). Coarse-grained RNA model for the martini 3 force field. *Biophysical J.* 125, 445–456. doi:10.1016/j.bpj.2025.07.034
- Yin, W., Zhang, Z., Zhang, S., He, L., Zhang, R., Jiang, R., et al. (2025). ERNIE-RNA: an RNA language model with structure-enhanced representations. *Nat. Commun.* 16, 10076. doi:10.1038/s41467-025-64972-0
- Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., et al. (2011). Biological validation that SF3b is a target of the antitumor macrocyclic pladienolide. *FEBS J.* 278, 4870–4880. doi:10.1111/j.1742-4658.2011.08387.x
- York, D. M. (2023). Modern alchemical free energy methods for drug discovery explained. *ACS Phys. Chem. Au* 3, 478–491. doi:10.1021/acspchemau.3c00033
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478, 64–69. doi:10.1038/nature10496
- Yoshida, K., Toki, T., Okuno, Y., Kanezaki, R., Shiraishi, Y., Sato-Otsubo, A., et al. (2013). The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat. Genet.* 45, 1293–1299. doi:10.1038/ng.2759
- Yoshida, H., Park, S.-Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., et al. (2015). A novel 3' splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes & Dev.* 29, 1649–1660. doi:10.1101/gad.267104.115
- Yoshida, H., Park, S.-Y., Sakashita, G., Nariai, Y., Kuwasako, K., Muto, Y., et al. (2020). Elucidation of the aberrant 3' splice site selection by cancer-associated mutations on the U2AF1. *Nat. Commun.* 11, 4744. doi:10.1038/s41467-020-18559-6
- Zhan, X., Lu, Y., and Shi, Y. (2024). Molecular basis for the activation of human spliceosome. *Nat. Commun.* 15, 6348. doi:10.1038/s41467-024-50785-0
- Zhang, Z., Chen, S., Chen, S., Chen, G., Zhang, R., Li, J., et al. (2017). SF3B1 mutation is a prognostic factor in chronic lymphocytic leukemia: a meta-analysis. *Oncotarget* 8, 69916–69923. doi:10.18632/oncotarget.19455
- Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. *Cell Res.* 28, 307–322. doi:10.1038/cr.2018.14
- Zhang, Z., Zhou, C., Tang, L., Gong, Y., Wei, Z., Zhang, G., et al. (2020). ASNEO: identification of personalized alternative splicing based neoantigens with RNA-seq. *Aging* 12, 14633–14648. doi:10.18632/aging.103516
- Zhang, Y., Qian, J., Gu, C., and Yang, Y. (2021). Alternative splicing and cancer: a systematic review. *Signal Transduct. Target. Ther.* 6, 78. doi:10.1038/s41392-021-00486-7
- Zhang, J., Huang, J., Xu, K., Xing, P., Huang, Y., Liu, Z., et al. (2022). DHX15 is involved in SUGP1-mediated RNA missplicing by mutant SF3B1 in cancer. *Proc. Natl. Acad. Sci.* 119, e2216712119. doi:10.1073/pnas.2216712119
- Zhang, Q., Ai, Y., and Abdel-Wahab, O. (2024). Molecular impact of mutations in RNA splicing factors in cancer. *Mol. Cell* 84, 3667–3680. doi:10.1016/j.molcel.2024.07.019
- Zhang, J., Yuan, R., Kryshchak, A., Kretsch, R. C., Schaeffer, R. D., Zhou, J., et al. (2025). Assessment of protein complex predictions in CASP16: are we making progress?. bioRxiv. 10.1101/2025.05.29.656875.
- Zhu, K., Trizio, E., Zhang, J., Hu, R., Jiang, L., Hou, T., et al. (2025). Enhanced sampling in the age of machine learning: Algorithms and applications. *Chem. Rev.* 126, 5c00700–5c00700713. doi:10.1021/acs.chemrev.5c00700