ChatGPT

CryoSTAR VAE Architecture and Latent Modeling

CryoSTAR: VAE with Structural Priors and Constraints

CryoSTAR is a recent method (2024) for cryo-EM heterogeneous reconstruction that employs a variational autoencoder (VAE) architecture guided by a reference atomic model [1]. In CryoSTAR's framework, an encoder network compresses each cryo-EM particle image into a continuous latent code, rather than assigning a discrete class label as in traditional 3D classification [2]. (For example, CryoSTAR uses an 8-dimensional latent space, and each particle image is mapped to a vector in this space representing its conformation.) The decoder then uses this latent vector to generate a **coarse-grained protein structure** – essentially an atomic model with residues positioned according to the latent variables – and can further produce a corresponding density map [1] [3]. This two-stage approach (first predicting atomic coordinates, then refining a density map) lets CryoSTAR output both an ensemble of atomic models and density volumes for different conformations [3].

Structural regularization: A key feature of CryoSTAR is the incorporation of **structural priors and constraints** from a given reference PDB model to regularize the VAE's learning [1]. In practice, the reference atomic model (docked into an initial consensus density) provides a starting conformation and a physical scaffold. CryoSTAR's decoder learns to *translate each residue* of this base structure independently according to the latent variables [4]. To preserve physically realistic motions, the model imposes additional losses that act as regularizers: (1) a *similarity loss* that penalizes large deviations between the deformed (predicted) structure and the reference structure (enforcing local rigidity), and (2) a *clash loss* to prevent unphysical steric overlaps between atoms [4]. These constraints serve as **atomic-level regularization** – essentially encouraging the latent space to generate plausible protein conformations that remain close to the reference geometry and avoid atomic clashes. CryoSTAR can also leverage simple chemical force-field terms (e.g. an elastic network of springs between nearby residues) to maintain chain connectivity and realism [4] [5]. By integrating these structural priors into the VAE loss function, CryoSTAR ensures that latent-space interpolations correspond to smooth, realistic conformational changes.

Latent distribution: In the vanilla CryoSTAR implementation, the VAE latent prior is presumably a standard Gaussian distribution (as is common in VAEs) – i.e. the model assumes an approximately normal latent space while training. Indeed, CryoSTAR's latent space is continuous and roughly Gaussian, which allows meaningful interpolation (they visualized principal components of the latent space to generate a series of conformations) [6] [7]. After training, the learned latent encodings of all particles can be analyzed: typically they form clusters or continuous trajectories corresponding to different conformational states. CryoSTAR's authors demonstrate this by performing PCA on the latent codes and sampling along principal axes, as well as clustering latent vectors to identify representative conformers [6] [2]. Notably, CryoSTAR's use of a continuous latent space (in contrast to rigid discrete classes) enables it to capture subtle gradations in conformation – an advantage for modeling continuous heterogeneity. However, using a simple Gaussian prior in the VAE can be a limiting factor if the true distribution of conformations is multi-modal or highly complex. CryoSTAR addresses part of this by the post-hoc clustering and by the structural regularization (which biases the latent distribution toward physically plausible regions), but the latent prior itself remains relatively simple (isotropic Gaussian) during training.

Could CryoSTAR Benefit from a Boltzmann Machine Latent Model?

Given the above, one might ask whether CryoSTAR's performance or latent representation could be improved by using a more expressive **Boltzmann machine-based prior or regularization** in the VAE. Boltzmann machines – such as Restricted Boltzmann Machines (RBMs) or Deep Boltzmann Machines (DBMs) – are energy-based models that can represent complex, multimodal distributions. Integrating a Boltzmann model into the VAE's latent space (in place of the simple Gaussian prior) could, in principle, allow the model to capture a more complicated landscape of protein conformations. For example, an RBM prior can learn a multi-peaked energy landscape in the latent space, which might naturally separate distinct conformational states by energy barriers (rather than forcing them into one Gaussian heap). This could help **preserve distinct conformers** and model their probabilities more faithfully, rather than blending them under a single Gaussian assumption.

Current research suggests that making the VAE prior more flexible in this way can indeed improve generative modeling. Boltzmann machine priors have been shown to be more expressive than factorized or Gaussian priors, able to encode non-Gaussian and correlated structure in the latent space [8]. In other words, a Boltzmann-type prior can capture complex latent dependencies and multimodality that a standard normal prior cannot. In the context of cryo-EM, this could mean better modeling of distinct energy minima corresponding to, say, "open" vs "closed" states of a protein, with fewer unrealistic intermediates. The potential benefit for CryoSTAR would be a latent space that more sharply distinguishes qualitatively different conformations (if they exist in the data) and perhaps a generative model less prone to producing unphysical blends of states. In high-noise regimes, a learned Boltzmann prior might also act as a stronger regularizer by encoding which latent regions correspond to viable structures, potentially improving robustness of reconstructions (similar in spirit to how CryoSTAR's structural prior guides the network).

However, there are practical challenges and trade-offs. Training a VAE with a Boltzmann machine prior is more complex – it typically requires running Markov Chain Monte Carlo (MCMC) sampling (e.g. Gibbs sampling) in the latent space during training to update the Boltzmann model, and the **evidence lower bound (ELBO)** no longer has a closed-form Kullback–Leibler term for the prior. Methods to integrate EBMs often involve approximate inference or additional neural estimators. This can slow down training and complicate convergence. CryoSTAR's current training already involves large datasets of images and a two-stage optimization; introducing an RBM/EBM would increase computation (and cryo-EM datasets might require a fairly large latent model to capture all variability). Moreover, CryoSTAR's structural constraints are a domain-specific regularization that might overlap with what a learned prior would provide. In essence, CryoSTAR already **imposes an informed structure on the generative model** (through the reference model and losses). Whether a Boltzmann prior would yield significant gains on top of that is an open question – it could, for instance, help if the data contain multiple distinct conformational basins that a single Gaussian latent cannot adequately represent. In summary, using a Boltzmann machine prior in CryoSTAR is theoretically appealing for richer latent modeling, but it comes with added complexity; the net benefit would depend on whether the conformational landscape demands a multimodal latent representation that current methods struggle to learn.

Recent VAE Architectures Integrating Boltzmann Machines

Several recent works (especially in the last couple of years) have explored **hybrid VAE models that incorporate Boltzmann machines or energy-based models in the latent space**. Below we summarize a few notable architectures and their contributions:

- **Discrete VAE with RBM Prior (RBM-VAE)** – *Gircha et al. (2023)*: This model introduces an RBM as the latent prior for a *discrete* variational autoencoder, applied to molecular generation [9] [10]. Instead of a continuous Gaussian, the latent variables are binary units with an RBM defining their joint distribution. The encoder produces a probability for each latent bit, and the RBM (with, e.g., 128 latent units) learns to "memorize" and model the latent patterns seen in data [11]. During training, RBM parameters are updated using contrastive divergence (Gibbs sampling in the latent space) to match the posterior distribution of the encoder [12] [13]. The contribution of this RBM-VAE is a more *flexible prior* that can capture multi-modal discrete latent features. In generative chemistry tasks, this allowed the model to generate **novel molecules** with valid chemical structures – the RBM prior helped constrain the latent space to regions corresponding to realistic molecules [14] [9]. Notably, the authors demonstrated training this hybrid on a quantum annealer as a proof-of-concept, since sampling from an RBM can be mapped to quantum hardware [15] [16]. The use of an RBM prior improved the diversity and validity of generated samples, showing that an energy-based latent can outperform simpler priors for complex data distributions.

- **DVAE++ and Extensions** – *Vahdat & Andriyash (2018)*: Earlier works laid the groundwork by integrating Boltzmann priors in VAEs. The DVAE (Discrete VAE) and its improved version DVAE++ used Boltzmann machines over binary latents to better capture discrete aspects of data [17]. These models showed that a learned Boltzmann prior can significantly improve the fidelity of reconstructions when the data has inherently discrete latent structure (e.g., class-like or clustering structure). Building on that idea, later research introduced **continuous relaxations** (e.g. DVAE#) to train VAEs with Boltzmann priors using tighter bounds and gradient methods [18] [19]. While DVAE++ is a bit older, we mention it because it established that *Boltzmann machines are effective priors for VAEs*, and recent work has continued along this direction.

- **Latent Energy-Based Models (EBM-VAEs)** – *Pang et al. (2020) and beyond*: An energy-based model (EBM) prior is a generalization of a Boltzmann machine to possibly continuous latent spaces. Pang et al. introduced a framework where the encoder network of a VAE is effectively replaced or augmented with an energy-based latent prior model (sometimes called a **LEBM**, latent EBM) [20]. Instead of assuming $p(z)\sim \mathcal{N}(0,I)$, they *learn* $p_\theta(z) \propto \exp(-E_\theta(z))$ with a neural energy function. Training this involves MCMC sampling of latent z during each iteration to estimate the prior's gradients. The result is a VAE that can learn much more complex latent distributions. Subsequent works have refined this approach – for example, recent studies propose improved training techniques for **VAEs with unnormalized priors**, addressing issues like the intractable partition function and optimization stability [21] [22]. The impact of EBM priors is seen in improved *latent structure and generative performance* [21]. One application in 2024 combined a latent EBM prior with diffusion models (LSD-EBM) for 3D medical image reconstruction, and found it outperformed a plain VAE in capturing anatomical details [23]. These EBM-VAEs highlight that using a learned energy landscape as the prior can yield higher-quality samples and a latent space that aligns better with data complexity, at the cost of more involved training.

- **Quantum Boltzmann VAE (QBM-VAE)** – *Wang et al. (2025)*: This cutting-edge approach embeds a **quantum Boltzmann machine** as the latent prior of a VAE [24] [8]. Essentially, it's an RBM-based VAE taken to the quantum hardware domain for efficient sampling. The motivation is that classical VAEs with Gaussian priors fail to capture the "complex, non-Gaussian landscapes" of real data (especially in biology) [25]. QBM-VAE uses discrete binary latents with an energy function $E(z)$ and leverages a quantum sampler (a coherent Ising machine/quantum annealer) to draw samples from the Boltzmann distribution defined by $E(z)$ [26] [27]. This hybrid quantum-classical VAE demonstrated **state-of-the-art performance on single-cell genomics data**, where the latent space learned by the Boltzmann prior preserved intricate structure in the data (e.g. cell developmental trajectories) far better than a standard Gaussian latent [8] [28]. In benchmarks on integrating multiple large single-cell datasets, the QBM-VAE outperformed conventional VAEs (and other probabilistic models like SCVI), achieving more faithful latent embeddings and more accurate downstream predictions [28]. The key insight is that a Boltzmann prior (even implemented via quantum means) brings *higher expressivity*: it can encode multimodal, highly correlated latent factors that match the complex reality of the data [8]. This example underscores the general point that **replacing a simplistic prior with a Boltzmann distribution can yield tangible improvements** in generative modeling of complex systems. (It also suggests that as computational tools improve, using such advanced priors might become more practical.)

- **Other Notable Variants**: Some works use **stacked RBMs or deep Boltzmann machines** as hierarchical priors for VAEs, though these are less common due to training difficulty. In principle, stacking RBMs (a DBM) could model even more complex latent hierarchies. Also related are VAE models with **Gaussian mixture priors** (not strictly Boltzmann machines, but another way to get a multimodal prior). For instance, a VAE with a Gaussian mixture prior can explicitly model clusters in latent space – conceptually simpler than an RBM, though less flexible than a full Boltzmann machine. Such mixture priors have been used to encourage the VAE to find discrete conformational states. While not the focus here, we mention it because CryoSTAR itself uses a Gaussian mixture *post hoc* to cluster latent codes; integrating a mixture prior during training could similarly enforce separated modes. Overall, the trend in recent years is toward **hybrid models** that combine VAEs with powerful latent models (EBMs, normalizing flows, etc.) to overcome the limitations of the plain Gaussian prior [22] [29]. The cited examples show that RBM/EBM-augmented VAEs have been successfully applied in diverse domains (anomaly detection in aeronautics, molecular design, medical imaging, single-cell biology), generally reporting better latent representations and generative accuracy than standard VAEs.

Latent Space Structure and Sampling: Boltzmann-VAEs vs Standard VAEs

Latent structure: The primary difference introduced by a Boltzmann machine prior is in the geometry and complexity of the learned latent space. In a standard VAE like CryoSTAR, the prior $p(z)$ is typically an isotropic Gaussian. This encourages the posterior latent codes to also distribute roughly in an ellipsoidal Gaussian cloud (the VAE training balances the reconstruction term with a KL-divergence pushing $q(z|x)$ towards $p(z)$). As a result, latent variables in CryoSTAR tend to form a single connected cluster (which can be probed by PCA, etc.), and transitions between different conformations in latent space are usually smooth interpolations [6] [2]. If the actual data has distinct conformations, the VAE posterior might still encode them along continuous directions within that one Gaussian ball – which can blur the boundaries between states. In contrast, **Boltzmann-augmented VAEs** can learn a latent space with multiple basins or more complicated topology. An RBM prior, for example, might learn that there are (say) two or three major modes in the binary latent unit space corresponding to fundamentally different structures. The latent posterior $q(z|x)$ will then be encouraged to place each

particle's code into one of those modes (or somewhere on a manifold shaped by the RBM's energy function). Thus, the latent space becomes *multi-modal or structured* in a way that aligns with the data distribution. Empirically, studies have found that VAE models with an RBM prior indeed capture latent clusters or multimodality that factorized priors miss [30] [31]. For example, Templin et al. (2023) observed that an RBM-based VAE could adapt to complex latent patterns and achieved better anomaly detection by modeling a richer latent distribution, whereas a simpler prior would force a poorer fit [31]. In summary, Boltzmann-based VAEs tend to yield latent spaces that *mirror the true data-generating factors more closely*: they can naturally represent multiple distinct states, sharp transitions or energy barriers between states, and non-linear correlations among latent dimensions.

Sampling and generation: In a standard VAE, sampling is straightforward – one draws $z \sim N(0,I)$ and feeds it to the decoder. Because the prior is simple, generating new samples or interpolating between known points is easy (though as noted, those samples might sometimes lie in regions the decoder wasn't well-trained on if the actual posterior was narrower or multimodal). With a Boltzmann machine prior, sampling becomes more involved: one must sample from the learned energy-based distribution $p_\theta(z)$. For an RBM, this typically means running Gibbs sampling or using an annealed sampler to draw binary configurations of z. For a continuous EBM prior, one might run Langevin dynamics or another MCMC method in latent space to obtain samples. These procedures are more computationally expensive and need to mix well to be reliable. Some advanced implementations use persistent chains (keeping a chain of latent samples that evolves during training) or specialized hardware (as in QBM-VAE using a quantum annealer) to speed this up [13] [32]. Once a latent sample is obtained, the decoder network generates the data as usual. The **generative performance** of Boltzmann-VAEs in practice has often been reported as superior to Gaussian-VAEs in terms of sample quality and data likelihood. For instance, the RBM-VAE for molecule generation was able to sample chemically valid molecules more often, because the RBM prior learned to put higher probability on latent codes corresponding to realistic chemistry [14] [9]. In another example, an EBM-VAE for images produced sharper, more accurate reconstructions than a standard VAE, since the flexible prior allowed the model to focus on probable image configurations rather than averaging over improbable ones [23]. That said, the improvement is not universal – if the data distribution is essentially unimodal or if the VAE posterior already matches a Gaussian fairly well, a more complex prior might not show big gains. In CryoSTAR's case, if the protein's conformational landscape is actually continuous (a single broad mode of variability), the Gaussian prior might be sufficient; but if there are multiple distinct substates, a Boltzmann prior could help segregate them in latent space.

In terms of **training stability**, introducing a Boltzmann prior can be tricky. VAEs with simple priors benefit from a convenient analytic KL term. With an unnormalized Boltzmann prior, one has to approximate the KL divergence (or train via alternative objectives). Techniques like variational inference with *unnormalized* priors have been developed to handle this, and researchers have compared different strategies to maintain stability and good convergence [21] [22]. Modern implementations often manage to train EBM-VAEs successfully, but care must be taken with hyperparameters (e.g., how long to run MCMC, how to initialize the sampler, temperature schedules, etc.). CryoSTAR's existing training pipeline would become more complex with these additions. Nonetheless, if done properly, the end result would be a model that perhaps better *shapes the latent space* according to the underlying physics of protein motions (one could imagine the Boltzmann prior learning an energy landscape akin to an actual free energy surface of the molecule's conformations).

Implications for CryoSTAR and Cryo-EM Modeling

For CryoSTAR's specific use case – learning the continuous heterogeneity of protein structures from cryo-EM images – incorporating Boltzmann machine techniques presents both potential benefits and challenges. On the upside, a Boltzmann/EBM-based latent model could enable **richer representation**

of conformational distributions. If a protein has a few distinct conformational states separated by energy barriers (for example, an open vs closed enzyme state, plus possibly some intermediate), a VAE with a flexible prior might naturally represent these as distinct modes in latent space (rather than compressing them into one cloudy latent cluster). This could improve the *identifiability* of states and allow the model to more cleanly generate representative structures for each state. In essence, it might reduce the "blurring" of conformations that sometimes occurs when a VAE tries to interpolate between states to satisfy a single Gaussian prior. A Boltzmann prior could act similarly to a Gaussian mixture or other structured prior in encouraging discrete groupings of latent codes, which aligns with how heterogeneity is often conceptualized (a few dominant conformers with transitions).

Another possible benefit is **better sampling of minor states**. CryoSTAR's current approach might under-represent low-probability conformations (since the VAE prior might pull encodings toward the dominant modes). An energy-based prior, however, can be shaped to have specific peaks – if the training data suggests a small subset of particles belong to an alternate conformation, the EBM could carve out a corresponding energy well in latent space. This could improve the model's ability to reconstruct and generate that minor state, as the prior would not overly penalize it. In cryo-EM analysis, capturing such rare conformations is valuable for understanding functional mechanisms.

Regularization vs. realism: CryoSTAR already uses strong domain-based regularization (the reference model constraints). A learned Boltzmann prior might reinforce some of those physical realism constraints by learning, for example, that latent combinations leading to clashes or chain breaks are simply low-probability (high energy). In fact, one could imagine training a Boltzmann prior that *itself* encodes knowledge of structural physics – e.g. an RBM could potentially learn patterns in latent space corresponding to concerted domain movements but penalize others. That is somewhat speculative, but energy-based models are in spirit related to physical energy landscapes. If one were to incorporate a Boltzmann prior in CryoSTAR, one could even initialize or bias it using biophysical insight (for instance, initialize an RBM's weights such that certain latent units correspond to motions of specific domains, etc., and let it refine from there). This synergy could further ensure generated structures are plausible.

On the downside, as discussed, adding such complexity may complicate training on large cryo-EM datasets. CryoSTAR's efficiency and effectiveness as reported [33] might be partly due to the simplicity of the latent prior which makes the inference easier. Introducing an RBM/EBM means introducing MCMC loops or additional neural networks to train (with many hyperparameters). There is also a risk of the Boltzmann prior learning something redundant or even conflicting with the structural prior: for example, if not carefully tuned, the learned energy function might pull the latent encodings in a way that the decoder's structural outputs start to violate the reference constraints (the model would then have to reconcile the learned prior with the fixed structural regularization). Ensuring that the two forms of regularization (one learned, one hard-coded) complement each other would be important.

Current research outlook: The idea of combining *physical priors with learned priors* is emerging in other domains. For cryo-EM specifically, methods like cryoSPHERE (2025) and DynaMight (2024) are exploring different ways to enforce physical realism in generative models of heterogeneity [5] [34]. These approaches do not yet use Boltzmann machines in latent space, but they share the goal of better conformational landscape modeling. It's conceivable that a future iteration of CryoSTAR or similar could experiment with a **Gaussian mixture prior or an EBM prior** to automatically partition heterogeneity into a few modes. The success of Boltzmann-augmented VAEs in other fields (as detailed above) is an encouraging sign. For instance, the QBM-VAE's ability to preserve subtle biological structure in latent embeddings [28] suggests that a learned prior can capture real-world complexity that a standard VAE might miss – an analogy in cryo-EM would be preserving subtle conformational differences in the latent space.

In conclusion, incorporating a Boltzmann machine-based latent model in CryoSTAR could *theoretically* enhance its capability to model complex conformational distributions by providing a more expressive prior. It might sharpen the separation of distinct states and improve generative accuracy for each state, potentially yielding more accurate and diverse atomic models of the protein's conformers. The trade-off is the added computational burden and modeling complexity. CryoSTAR's current design already addresses heterogeneity with the clever use of a reference structure and VAE – a Boltzmann prior would be a next-level improvement aimed at capturing *probabilistic structure* in the latent space itself. Whether this yields a tangible improvement for cryo-EM datasets would need to be validated experimentally. Given the trends in recent VAE research and the analogous successes in other domains, it is a promising direction: a Boltzmann-augmented CryoSTAR could, in principle, marry data-driven flexibility with physical realism even more closely, potentially advancing the state-of-the-art in cryo-EM structural ensemble modeling.

Sources:

- CryoSTAR methodology and VAE architecture [1] [2] [4]
- CryoSTAR outputs and latent space analysis [3] [6]
- CryoSTAR regularization using reference model (similarity and clash losses) [4]
- Benefits of RBM/energy-based priors in VAEs (anomaly detection and flexibility) [31] [35]
- Latent energy-based models (EBM prior) for VAEs [20] [23]
- RBM-VAE and DVAE++ contributions [17] [14]
- Quantum Boltzmann VAE performance on complex data [8] [28]
- QBM-VAE (RBM prior) outperforming Gaussian VAE in single-cell integration [36]

---

[1] [3] [33] cryoSTAR
https://bytedance.github.io/cryostar/

[2] [6] [7] A Minimal Case | CryoSTAR
https://byte-research.gitbook.io/cryostar/a-minimal-case

[4] [5] [34] cryoSPHERE: Single-Particle HEterogeneous REconstruction from cryo EM
https://arxiv.org/html/2407.01574v2

[8] [24] [26] [27] [28] Quantum Boltzmann Machine-VAE
https://www.emergentmind.com/topics/quantum-boltzmann-machine-variational-autoencoder-qbm-vae

[9] [10] [11] [12] [13] [14] [15] [16] Hybrid quantum-classical machine learning for generative chemistry and drug design | Scientific Reports
https://www.nature.com/articles/s41598-023-32703-4?error=cookies_not_supported&code=b436034e-57b5-419d-b703-f4c15581b223

[17] [18] [19] DVAE#: Discrete Variational Autoencoders with Relaxed Boltzmann Priors
http://papers.neurips.cc/paper/7457-dvae-discrete-variational-autoencoders-with-relaxed-boltzmann-priors.pdf

[20] [23] Energy-Based Prior Latent Space Diffusion model for Reconstruction of Lumbar Vertebrae from Thick Slice MRI
https://arxiv.org/html/2412.00511v1

[21] [22] [29] openreview.net
https://openreview.net/pdf/3491728fa1b8d36a0dcc90bec4fe189cd68baeb2.pdf

[25] [2508.11190] Quantum-Boosted High-Fidelity Deep Learning
https://arxiv.org/abs/2508.11190

30 31 35 Performance of VAE models with Gaussian, Bernoulli, and RBM priors in... | Download Scientific Diagram
https://www.researchgate.net/figure/Performance-of-VAE-models-with-Gaussian-Bernoulli-and-RBM-priors-in-the-baseline-study_fig5_372356611

32 36 Quantum Boltzmann-Variational Autoencoder Improves Deep Generative Modelling Of Complex Data
https://quantumzeitgeist.com/quantum-boltzmann-variational-autoencoder-improves-deep-generative-modelling-of-complex-data/