GNMCADS: SAMPLING FOR PROTEIN CONFORMATION DIVERSITY WITH GAUSSIAN NETWORK MODEL BASED CONDITION ANNEALED DIFFUSION SAMPLER

Anonymous authors
Paper under double-blind review

ABSTRACT

Proteins are dynamic molecules existing in multiple conformational states that dictate their biological functions. Current methods for predicting protein conformational ensembles mostly rely on multiple sequence alignment (MSA) modifications to increase the diversity of single-state predicting models, or incorporate molecular dynamics (MD) simulations during training to mimic conformational variability. However, MSA modifications can be insufficient for capturing conformational changes, particularly in proteins with low sequence homology such as *de novo* designed proteins and in protein complexes, where MSAs are constructed on a per-chain basis. Similarly, MD trained methods can fail to capture the conformational changes that happen in large timescales or large protein complexes as it is infeasible to perform MD simulations for these cases. Here, we propose a new diffusion sampling strategy that can be applied to any diffusion protein generative model that uses pairwise conditioning signals without any additional training. Our method GNMCADS, enhances structural diversity by increasing the uncertainty between the dynamical domains derived by the Gaussian Network Model (GNM), using Condition Annealed Diffusion Sampler (CADS). As GNM naturally extends to multi chain systems and does not depend on MSA, GNMCADS enables the sampling of long timescale conformational changes of proteins, including those with low evolutionary informations and of multimeric complexes. We further implement GNMCADS on AlphaFold3 (AF3) and compare it with current state-of-the-art MSA modification or MD based conformational sampling methods in 38 cases.

1 INTRODUCTION

Proteins are the essential machines that regulate and execute cellular activities. With the advent of deep learning methods such as AlphaFold2 (Jumper et al., 2021), and its successor AlphaFold3 (AF3) (Abramson et al., 2024), it is now possible to accurately predict a protein's 3 dimensional structure from its sequence. However, proteins are highly dynamic molecules and exist in multiple conformational states that govern their functions. Therefore, understanding the conformational landscapes of proteins is essential for understanding their functions (Henzler-Wildman & Kern, 2007).

Currently, there exists two main approaches for predicting the conformational landscape of a protein. The first approach leverages the fact that AlphaFold uses multiple sequence alignments (MSA) to infer co-evolutionary information about the given protein. Consequently, modifications of MSA can improve the diversity of the generated outputs (Wayment-Steele et al., 2024; Kalakoti & Wallner, 2025; Del Alamo et al., 2022; Stein & Mchaourab, 2022; Passaro et al., 2025). However, as the MSA is calculated for each chain individually, MSA modification based methods struggle to generate conformational landscapes in multi chain structures where the conformational change depends on the chain interactions. Furthermore, as MSA is an optional input for AlphaFold, modifications of MSA can be insufficient to predict the conformational of especially proteins with low sequence homology such as *de novo* designed proteins therefore low aligned sequence count. The second approach involves training deep learning models with structures from molecular dynamics (MD) simulations (Jing et al., 2024; Passaro et al., 2025; Lewis et al., 2025; Zheng et al., 2024). However, even with the help of enhanced sampling techniques it is infeasibly expensive to compute MD simulations for

large timescales or for large protein complexes. Hence, the models trained with MD simulations can fail to predict conformational states with large timescale changes and conformational landscapes of protein complexes.

The challenge of sampling diverse outputs is not unique to the protein generative models. It has been shown possible to increase the diversity of diffusion models by techniques such as guidance mechanisms (Dhariwal & Nichol, 2021; Ho & Salimans, 2022; Hong et al., 2022), adding noise to the cascading inputs of cascading diffusion models (Ho et al., 2021), or new sampling strategies such as Condition Annealed Diffusion Sampler (CADS) (Sadat et al., 2023). However as it can be seen in Appendix D.1, these techniques can not be trivially applied to the protein generative models.

We introduce Gaussian Network Model Condition Annealed Diffusion Sampler (GNMCADS), which combines the Gaussian Network Model's (GNM) (Haliloglu et al., 1997; Bahar et al., 1997) ability to predict the equilibrium dynamical properties and the dynamical domains of proteins with the CADS' ability to increase the diversity of diffusion models to sample physically grounded diverse protein conformations by increasing the uncertainty of the diffusion conditioning signal selectively between the dynamical domains of the sampled protein.

As GNMCADS is a sampling strategy, it does not require retraining and can be applied to any protein diffusion generative model that uses residue pair conditioning signals. Here, we implement GNMCADS on AlphaFold3's diffusion module and compare its ability to generate multiple conformations with AFSample2 (Kalakoti & Wallner, 2025) which randomly masks MSA columns to increase output diversity, Boltz-2 (Passaro et al., 2025) which is the open-source implementation of the AlphaFold3 that is also trained with MD simulations and can generate multiple conformations by mimicking molecular dynamics or sub sampling MSA, and BioEmu (Lewis et al., 2025) which is a generative model that emulates equilibrium distribution of coarse grained proteins.

2 METHOD

2.1 GAUSSIAN NETWORK MODEL

The Gaussian Network Model (GNM) is a physics-based, a minimalist coarse-grained normal mode model that describes the equilibrium dynamics of proteins (Haliloglu et al., 1997; Bahar et al., 1997). In GNM, a protein structure is represented as an elastic network in which C^{α} atoms that lie within a cutoff distance are connected by uniform harmonic springs. Under the assumption of Gaussian-distributed fluctuations, this representation enables the characterization of equilibrium motions within a statistical thermodynamics framework, where collective residue fluctuations are decomposed into intrinsic normal modes of the network. Further details about GNM can be found in Appendix A.

GNM provides dynamical cross-correlations between residue motions. For a given dynamic mode, the sign of the cross-correlation between two residues indicates whether they tend to move in the same or opposite directions. As a result, each dynamic mode partitions the protein into two dynamical domains, such that residues within the same domain move coherently, whereas residues belonging to opposing domains exhibit anticorrelated motions. The dynamical domains of protein O76728 associated with the two slowest (lowest-frequency) modes are shown in Figure 1a.

2.2 GAUSSIAN NETWORK MODEL BASED CONDITION ANNEALED SAMPLER (GNMCADS)

GNMCADS works by annealing the conditioning signal in a physically grounded way and can be used in any protein diffusion model that includes pairwise representation as a conditioning signal such as AlphaFold3. The pairwise conditioning signal is only annealed by adding Gaussian noise to the inter dynamical domain regions calculated by GNM, as can be seen in Figure 1a.

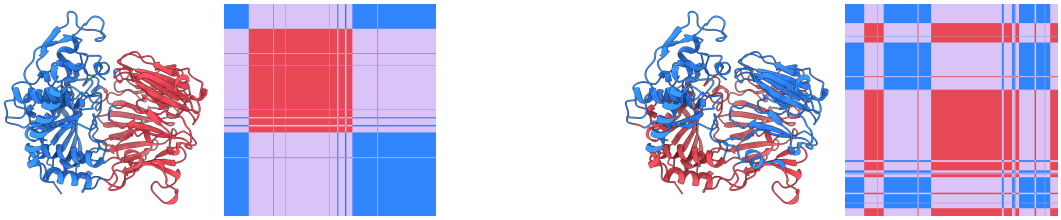
Low-frequency GNM modes encode the collective functional motions (Bahar et al., 2010; Haliloglu & Bahar, 2015; Aykac Fas et al., 2025). Many experimentally observed conformational transitions can thus be well described by the lowest frequency GNM modes. Accordingly, at each diffusion step, the mode of motion used to define dynamical domains is selected from a distribution that favors low-frequency modes obtained by their eigenvalues.

As it has been shown that addition of noise to the diagonals of pair representation in AlphaFold3 like models can cause significant effects in the secondary structure (Cho et al., 2025), GNMCADS also masks the diagonal of conditioning signal to preserve the secondary structure.

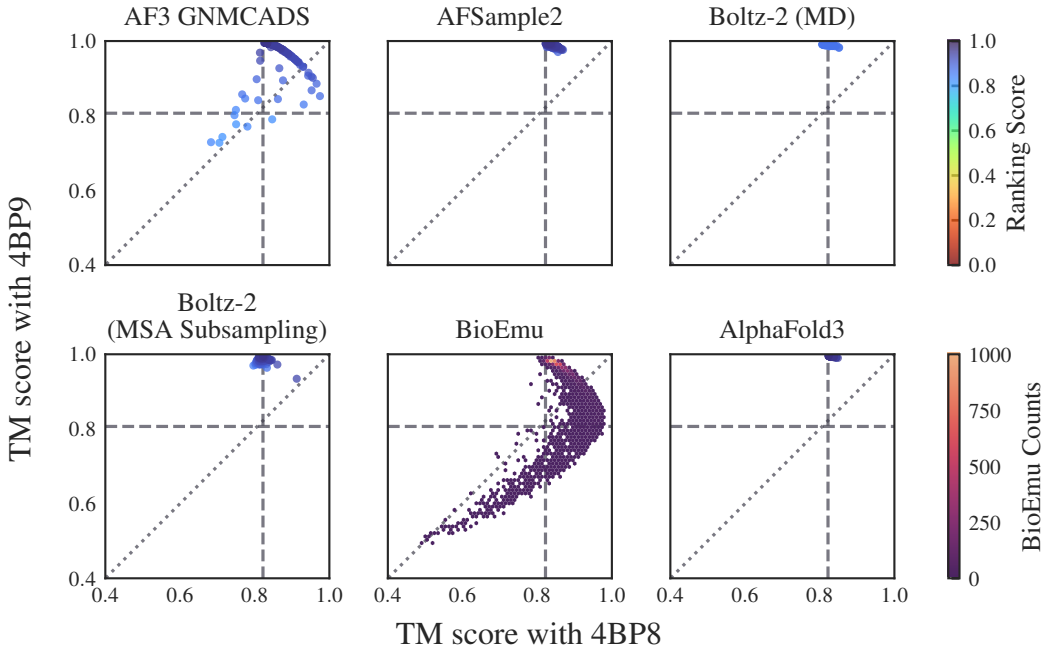
Similar to CADS, GNMCADS is also performed with a linear annealing schedule. However, GNMCADS do not use rescaling as the addition of noise in each step of the diffusion inference is not uniform. Further details about the effects of hyperparameters can be found in Appendix D.

Since the dynamical domain calculation with GNM requires a structures, the first structure is sampled without annealing the conditioning signal to be used in GNM calculations. Further details on the implementation of GNMCADS on AlphaFold3 (AF3 GNMCADS) can be found in Appendix B.

3 RESULTS



(a) AF3 predicted structures and residue pair matrices of O76728 colored by the dynamical domains calculated with GNM. Residue pair matrices include the dynamical tendency to *Left*: Dynamical modes of the slowest GNM mode. *Right*: Dynamical modes of the second slowest GNM mode. *Blue*: Dynamical domain A. *Red*: Dynamical domain B. *Pink*: Inter-domain region.



(b) TM similarity scores of the predicted structures of O76728 by AF3 GNMCADS, AFSample2, Boltz-2 (MD), Boltz-2 (MSA Subsampling), BioEmu, and AlphaFold3. Structures are colored by their predicted ranking scores except BioEmu. As BioEmu does not predict a ranking score, its results are shown with a hex bin plot. Dashed lines represent the TM similarity scores between the PDB structures 4BP9 and 4BP8. Dotted line represent the $y = x$ line. Only structures with higher ranking score than 0.8067, the TM similarity between 4BP9 and 4BP8, is shown. See Appendix C for further details and comparisons.

Figure 1: Conformational landscape sampling and GNM dynamical modes of the protein O76728.

The ability of AF3 GNMCADS, AFSample2, Boltz-2 (MD), Boltz-2 (MSA Subsampling) and BioEmu to predict heterogeneous protein conformations was benchmarked on the Open-Closed dataset (Kalakoti & Wallner, 2025) and the Transporters dataset (Xie & Huang, 2024). As it can be seen from the Figures 1b and 2, AF3 GNMCADS can successfully predict a diverse set of conformations for proteins with multiple conformations. Additional details and diversity plots for other targets in the datasets can be found in Appendix C.

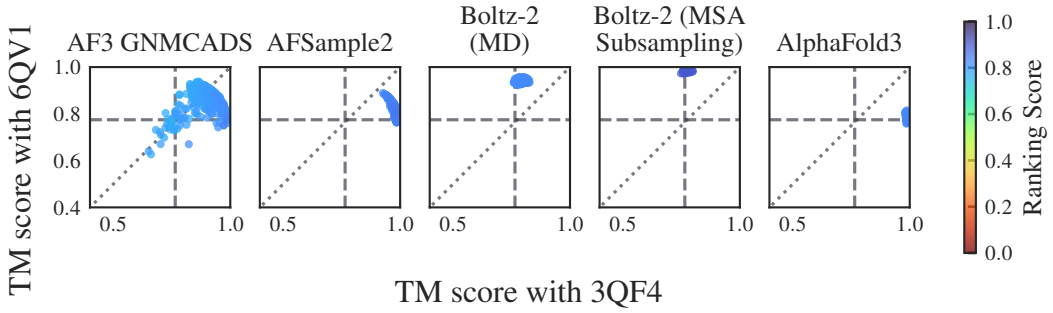
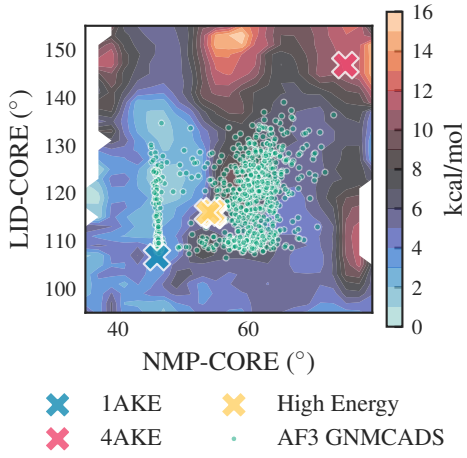
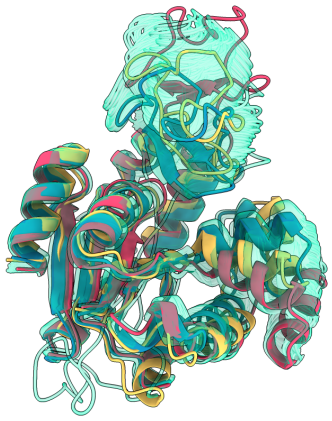


Figure 2: Diversity plots of the predicted structures of TM_0287 by AF3 GNMCADS, AFSample2, Boltz-2 (MD), Boltz-2 (MSA Subsampling), and AlphaFold3. Structures are colored by their predicted ranking scores. BioEmu is not shown as it does not support structures with multiple chains. Dashed lines represent the TM similarity scores between the PDB structures 3QF4 and 6QV1. Dotted line represent the $y = x$ line. Only structures with higher ranking score than 0.7646, the TM similarity between 3QF4 and 6QV1, is shown. See Appendix C for further details and comparisons.



(a) AF3 GNMCADS predicted conformational ensemble of *E. Coli* adenylate kinase projected onto its free energy landscape defined by two collective variables describing the global orientations of its domains, together with the experimental structures 1AKE and 4AKE, and high energy NMR structures of *Geobacillus sp.* adenylate kinase.



(b) AF3 GNMCADS predicted structures of *E. Coli* adenylate kinase (green) aligned to PDB structures 1AKE (blue), 4AKE (red), and high energy NMR structures of of *Geobacillus sp.* adenylate kinase (yellow).

Figure 3: The conformations predicted by AF3 GNMCADS match against experimentally determined high energy structures and free energy landscape.

Building on these benchmarking results, the conformational diversity of *E. Coli* adenylate kinase captured by AF3 GNMCADS are matched with the free energy landscape obtained by enhanced MD simulations (Beckstein et al., 2009). Similarly, the predicted conformations are consistent with experimentally determined high-energy adenylate kinase structures from *Geobacillus sp.* (Stiller et al., 2022).

The effect of the protein dynamics is also evaluated by performing predictions with AF3 GNMCADS where the GNM analysis is carried on a unique freely jointed chain for each target where further details can be found in Appendix D.2.

4 Discussion And Conclusion

In this work, we introduce GNMCADS, a diffusion sampling strategy that combines the Gaussian Network Models with Condition Annealed Diffusion Sampling to reveal the diverse set of conformations of diffusion based protein generative models for single and multi chain structures without depending on multiple sequence alignments or needing retraining with molecular dynamic simulation data. Consequently, we integrate GNMCADS on AlphaFold3's diffusion module, assessed its ability to generate conformations on 38 different protein targets and compared it with other state-of-the-art MSA modification and MD simulation based methods that aim to generate diverse protein conformations. We demonstrate that by annealing the diffusion condition signal using CADS guided by the intrinsic dynamics of the protein of interest calculated by GNM can be used to successfully sample the diverse conformations. Further, since the Gaussian Network Model can be used to analyze the equilibrium dynamics of protein nucleic acid complexes, GNMCADS in principle also be applied to enhance the sampling of such assemblies.

Contrary to GNMCADS, conformation sampling methods trained on MD simulations such as Boltz-2 (MD) or BioEmu demand a large amount of compute power both for the generation of MD trajectories and for the training process. Moreover, combining the ability of generating diverse conformations of these models with other existing generative models is a non-trivial challenge whereas GNMCADS can be integrated readily to proteins diffusion models that use pairwise conditioning.

Since the GNMCADS only modifies the diffusion sampling process, compared to MSA modification based methods, which require to restart the inference process for each MSA modification, GNMCADS can sample a diverse set of conformations by annealing the same conditioning signal. Hence, for AlphaFold-like protein generative models where the main computational bottleneck lies in creating the conditioning signal such as in the pairformer module of AlphaFold3, GNMCADS only requires a single inference pass that generates the diffusion conditioning signal. Likewise, as MSA used as an input to other tasks of protein generative models such as confidence prediction modules of AlphaFold or ligand affinity modules of Boltz-2, modifications of MSA may result in unintended side effects across other modules that depend on MSA. In addition, MSA modifications can be ineffective for proteins with low sequence homology such as *de novo* designed binders.

Together, these results position GNMCADS as a computationally efficient and broadly applicable framework for enhancing conformational diversity in diffusion based protein generates models.

References

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07487-w.

Burcu Aykac Fas, Zeynep Erge Akbas Buz, and Turkan Haliloglu. Global dynamics behind enzyme catalysis, evolution, and design. *Current Opinion in Structural Biology*, 94:103131, October 2025. ISSN 0959440X. doi: 10.1016/j.sbi.2025.103131.

Ivet Bahar, Ali Rana Atilgan, and Burak Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173–181, June 1997. ISSN 13590278. doi: 10.1016/S1359-0278(97)00024-2.

Ivet Bahar, Timothy R. Lezon, Lee-Wei Yang, and Eran Eyal. Global Dynamics of Proteins: Bridging Between Structure and Function. *Annual Review of Biophysics*, 39(1):23–42, April 2010. ISSN 1936-122X, 1936-1238. doi: 10.1146/annurev.biophys.093008.131258.

Oliver Beckstein, Elizabeth J. Denning, Juan R. Perilla, and Thomas B. Woolf. Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open $\leftrightarrow$ Closed Transitions. *Journal of Molecular Biology*, 394(1):160–176, November 2009. ISSN 00222836. doi: 10.1016/j.jmb.2009.09.009.

Yehlin Cho, Griffin Rangel, Gaurav Bhardwaj, and Sergey Ovchinnikov. Protein Hunter: Exploiting structure hallucination within diffusion for protein design, October 2025.

Diego Del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, 11:e75751, March 2022. ISSN 2050-084X. doi: 10.7554/eLife.75751.

Prafulla Dhariwal and Alex Nichol. Diffusion Models Beat GANs on Image Synthesis, 2021.

Turkan Haliloglu and Ivet Bahar. Adaptability of protein structures to enable functional interactions and evolutionary implications. *Current Opinion in Structural Biology*, 35:17–23, December 2015. ISSN 0959440X. doi: 10.1016/j.sbi.2015.07.007.

Turkan Haliloglu, Ivet Bahar, and Burak Erman. Gaussian Dynamics of Folded Proteins. *Physical Review Letters*, 79(16):3090–3093, October 1997. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.79.3090.

Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450 (7172):964–972, December 2007. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature06522.

Jonathan Ho and Tim Salimans. Classifier-Free Diffusion Guidance, 2022.

Jonathan Ho, Chitwan Saharia, William Chan, David J. Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded Diffusion Models for High Fidelity Image Generation, 2021.

Susung Hong, Gyuseong Lee, Wooseok Jang, and Seungryong Kim. Improving Sample Quality of Diffusion Models Using Self-Attention Guidance, 2022.

Bowen Jing, Bonnie Berger, and Tommi Jaakkola. AlphaFold Meets Flow Matching for Generating Protein Ensembles, 2024.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03819-2.

Yogesh Kalakoti and Björn Wallner. AFsample2 predicts multiple conformations and ensembles with AlphaFold2. *Communications Biology*, 8(1):373, March 2025. ISSN 2399-3642. doi: 10.1038/s42003-025-07791-9.

Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew Y. K. Foong, Victor García Satorras, Osama Abdin, Bastiaan S. Veeling, Iryna Zaporozhets, Yaoyi Chen, Soojung Yang, Adam E. Foster, Arne Schneuing, Jigyasa Nigam, Federico Barbero, Vincent Stimper, Andrew Campbell, Jason Yim, Marten Lienen, Yu Shi, Shuxin Zheng, Hannes Schulz, Usman Munir, Roberto Sordillo, Ryota Tomioka, Cecilia Clementi, and Frank Noé. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761):eadv9817, August 2025. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adv9817.

Elaine C. Meng, Thomas D. Goddard, Eric F. Pettersen, Greg S. Couch, Zach J. Pearson, John H. Morris, and Thomas E. Ferrin. UCSF CHIMERAX : Tools for structure building and analysis. *Protein Science*, 32(11):e4792, November 2023. ISSN 0961-8368, 1469-896X. doi: 10.1002/pro.4792.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv*, pp. 2025.06.14.659707, January 2025. doi: 10.1101/2025.06.14.659707.

Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann M. Weber. CADS: Unleashing the Diversity of Diffusion Models through Condition-Annealed Sampling. 2023. doi: 10.48550/ARXIV.2310.17347.

Richard A. Stein and Hassane S. Mchaourab. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLOS Computational Biology*, 18(8):e1010483, August 2022. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1010483.

John B. Stiller, Renee Otten, Daniel Häussinger, Pascal S. Rieder, Douglas L. Theobald, and Dorothee Kern. Structure determination of high-energy states in a dynamic protein ensemble. *Nature*, 603(7901):528–535, March 2022. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-022-04468-9.

Michael L. Waskom. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6 (60):3021, 2021. doi: 10.21105/joss.03021.

Hannah K. Wayment-Steele, Adedolapo Ojoawo, Renee Otten, Julia M. Apitz, Warintra Pitsawong, Marc Hömberger, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature*, 625(7996):832–839, January 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-023-06832-9.

Tengyu Xie and Jing Huang. Can Protein Structure Prediction Methods Capture Alternative Conformations of Membrane Transporters? *Journal of Chemical Information and Modeling*, 64(8): 3524–3536, April 2024. ISSN 1549-9596, 1549-960X. doi: 10.1021/acs.jcim.3c01936.

Sandra Zakrzewska, Ahmad Reza Mehdipour, Viveka Nand Malviya, Tsuyoshi Nonaka, Juergen Koepke, Cornelia Muenke, Winfried Hausner, Gerhard Hummer, Schara Safarian, and Hartmut Michel. Inward-facing conformation of a multidrug resistance MATE family transporter. *Proceedings of the National Academy of Sciences*, 116(25):12275–12284, June 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1904210116.

Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. US-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods*, 19(9): 1109–1115, September 2022. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-022-01585-1.

Shuxin Zheng, Jiyan He, Chang Liu, Yu Shi, Ziheng Lu, Weitao Feng, Fusong Ju, Jiaxi Wang, Jianwei Zhu, Yaosen Min, He Zhang, Shidi Tang, Hongxia Hao, Peiran Jin, Chi Chen, Frank Noé, Haiguang Liu, and Tie-Yan Liu. Predicting equilibrium distributions for molecular systems with deep learning. *Nature Machine Intelligence*, 6(5):558–567, May 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00837-3.

A BACKGROUND ON THE GAUSSIAN NETWORK MODEL

The Gaussian Network Model (GNM) represents a protein as an elastic network in which C^{α} atoms within a cutoff distance $r_c = 10\text{Å}$ are connected by uniform harmonic springs undergoing Gaussian-distributed fluctuations. This formulation enables the equilibrium cross-correlations of fluctuations between residues i and j to be expressed as

$$\langle \Delta \boldsymbol{R_i} \cdot \Delta \boldsymbol{R_j} \rangle = (k_B T/\gamma)[\boldsymbol{\Gamma^{-1}}]_{\boldsymbol{ij}} \tag{1}$$

where $\boldsymbol{R_i}$ is the position vector of the ith residue's C^{α} atom, k_B is the Boltzmann constant, T is the absolute temperature, γ is the force constant, and $\mathbf{\Gamma}$ is the Kirchhoff matrix representing the connectivity of the protein such that

$$\mathbf{\Gamma_{ij}} = \begin{cases} -1 & i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{k,k\neq j} \mathbf{\Gamma_{kj}} & i = j \end{cases} \tag{2}$$

where R_{ij} is the distance between the C^{α} atoms of the residues i and j.

The pseudo-inverse of $\mathbf{\Gamma^{-1}}$ can be represented by the sum of the $N-1$ non-zero modes $2 \leq k \leq N$ using

$$\mathbf{\Gamma^{-1}} = \sum_{k=2}^{N} \lambda_k^{-1} \boldsymbol{u_k} \boldsymbol{u_k^T} \tag{3}$$

where λ_k and $\boldsymbol{u_k}$ are the eigenpairs of $\mathbf{\Gamma}$. The first mode $k = 1$ is omitted as it has an eigenvalue of 0.

Since $\langle \Delta \boldsymbol{R_i} \cdot \Delta \boldsymbol{R_j} \rangle$ is directly proportional to $\mathbf{\Gamma_{ij}^{-1}}$, the correlations of the equilibrium fluctuations between the residues i and j for the mode k can be calculated as

$$C_{ij}^{(k)} = \frac{1}{\lambda_k} \boldsymbol{u_k}(i) \boldsymbol{u_k}(j) \tag{4}$$

B Implementation of GNMCADS on AlphaFold3

The GNMCADS diffusion sampling strategy can be implemented to any diffusion based protein generative model that uses pair conditioning signals such as AlphaFold3. Implementation of GNMCADS to AlphaFold3 has been performed by annealing the pairwise representation that has been computed by the trunk region of AlphaFold3.

Since to analyze the dynamical modes with GNM requires an input structure, first 5 structures are sampled without condition annealing. The structure with highest predicted ranking score is used to determine the dynamical cross-correlations between residues using GNM to define the dynamical domains for each modes of motion. Consequently, the distribution that will be used to select the mode of motion to be used in each diffusion step is created by performing the softmax operation on the negative eigenvalues of modes to favour the low frequency modes.

The pairwise conditioning signal is only corrupted for inter-domain regions where the residues i and j belong to different dynamical domains.

$$\widehat{\boldsymbol{z_{ij}^{\text{trunk}}}} = \begin{cases} \sqrt{\gamma(t)} \boldsymbol{z_{ij}^{\text{trunk}}} + s\sqrt{1-\gamma(t)} \boldsymbol{n_{ij}} & \text{domain}_i \neq \text{domain}_j, \\ \boldsymbol{z_{ij}^{\text{trunk}}} & \text{domain}_i = \text{domain}_j \end{cases} \tag{5}$$

where the noise scale s is selected as 0.3, $\boldsymbol{n_{ij}}$ is sampled from a Gaussian distribution such that $\boldsymbol{n_{ij}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, and $\gamma(t)$ is the linear annealing schedule such that

$$\gamma(t) = \begin{cases} 1 & t \leq \tau_1, \\ \frac{\tau_2 - t}{\tau_2 - \tau_1} & \tau_1 < t < \tau_2, \\ 0 & t \geq \tau_2, \end{cases} \tag{6}$$

where the values of τ_1 and τ_2 is selected as $\tau_1 = 0$ and $\tau_2 = 0.3$.

C Comparisons with other conformational landscape sampling methods

The ability to predict diverse conformational samples of the methods have been evaluated on the Open-Closed dataset proposed by Kalakoti & Wallner (2025) and Transporters dataset proposed

Table 1: Open-Closed dataset targets

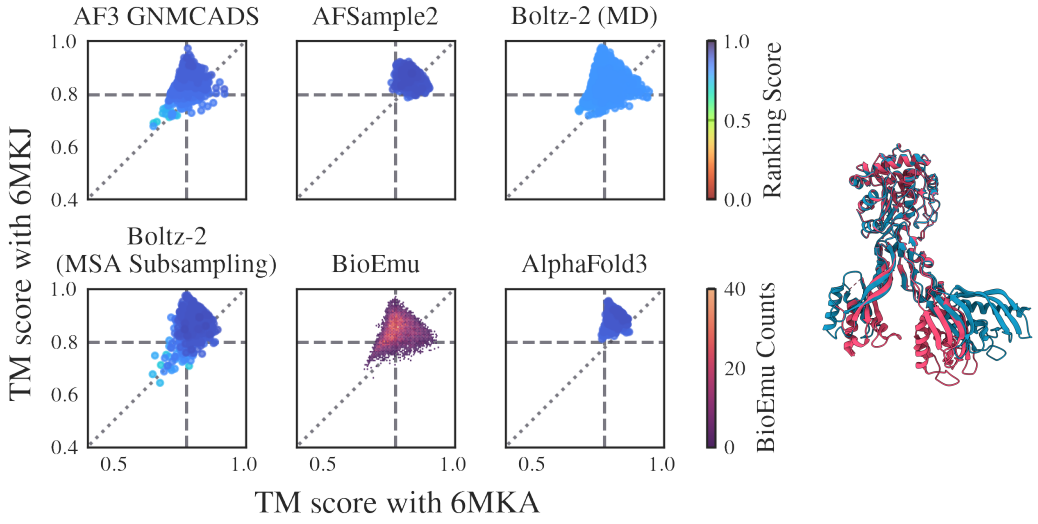
Target Name (UniProt ID)	Conformation 1 PDB ID	Conformation 2 PDB ID
A0A075Q0W3	6MKA	6MKJ
A0QTT2	7CY2	7CYR
A2RJ53	3FTO	3DRF
A6UVT1	6HAC	6HAE
B3EYN2	5HO2	5HO0
B7IE18	6NC7	6NC6
O76728	4BP8	4BP9
P00558	2XE6	2WZD
P21589	7QGA	4H2I
P31133	6YED	6YE0
P33284	3O6W	3O8M
P40131	3TEE	3VJP
P62495	2KTV	2KTU
P71447	2WFA	2WF5
Q18A65	6HNJ	6HNI
Q53W80	7C63	7C6N
Q53W80	7C67	7C64
Q5F9M1	3ZSF	2YLN
Q7DAU8	3L6G	3L6H
Q9ERE7	2RQM	2RQK
Q9SS90	6K8B	6K85
Q9X6R4	3IUJ	3IUQ
Q9X9P9	2OLO	2OLN
Q9Z4N6	1SI1	1SI0

by Xie & Huang (2024). The AF3 GNMCADS predictions were performed with only one seed and 1000 diffusion sampled protein structures. The structures sampling with AFSample2 were performed with 1000 samples. However, for multi chain structures 1005 samples were predicted to ensure the equal number of structures generated by each AlphaFold2 model that AFSample2 uses. Similarly, the Boltz-2 MD predictions were performed by sampling 1000 protein structures. The Boltz-2 MSA subsampling predictions were performed by sampling 50 protein structures for each MSA depth of of 16, 32, 64, 128, 256, 512, 1024, and 5120 as recommended by Del Alamo et al. (2022). The BioEmu predictions were performed by sampling 10,000 protein structures as it has been noted on Lewis et al. (2025). The AlphaFold3 predictions were performed with 1000 seeds with 5 predictions per seed.

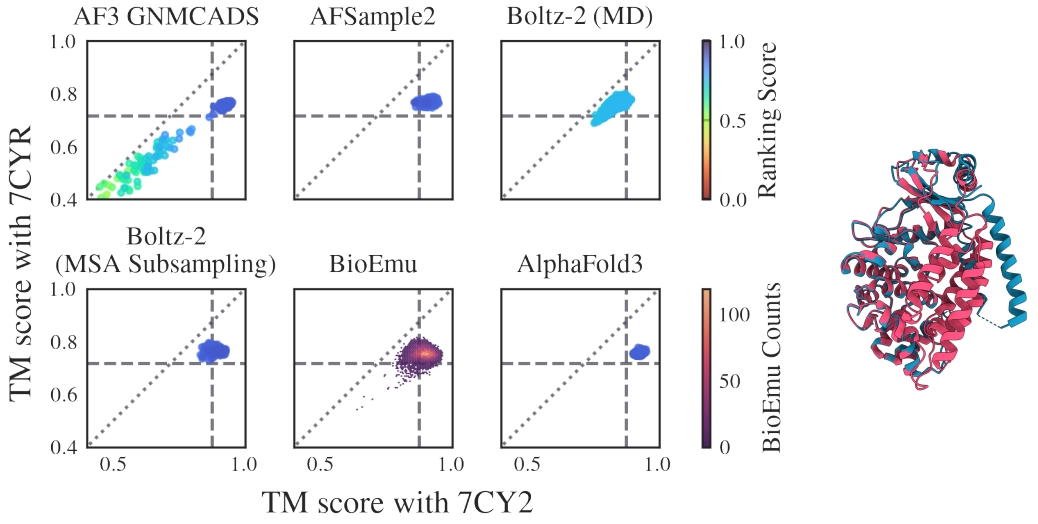
The similarity calculations between proteins were performed using US-align (Zhang et al., 2022). The diversity plots were created with seaborn (Waskom, 2021), whereas the protein structure visualizations were conducted with UCSF ChimeraX (Meng et al., 2023).

C.1 Open-Closed dataset

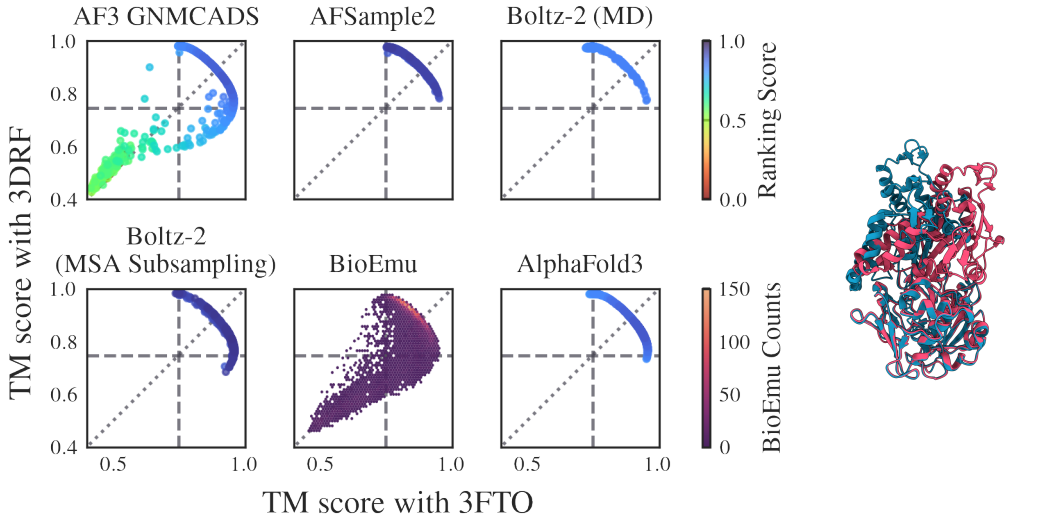
The Open-Closed dataset contains 23 protein targets with 24 different conformational pairs. In contrast to the version proposed by Kalakoti & Wallner (2025), predictions were performed with all of the protein chains included in the first biological assembly for each target. As the protein Q53W80 also exists as a homodimer, it is divided into two targets for its monomer and dimer form. The PDB Identifiers for each Open-Closed dataset target can be seen in Table 1.



(a) *Left*: Diversity plots of A0A075Q0W3. *Right*: Structures 6MKA (blue), and 6MKJ (red).

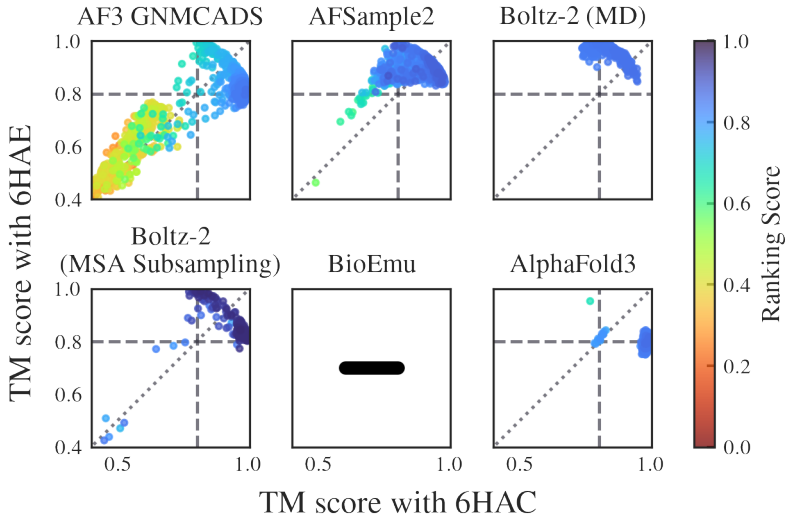


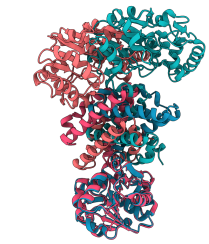
(b) *Left*: Diversity plots of A0QTT2. *Right*: Structures 7CY2 (blue), and 7CYR (red).



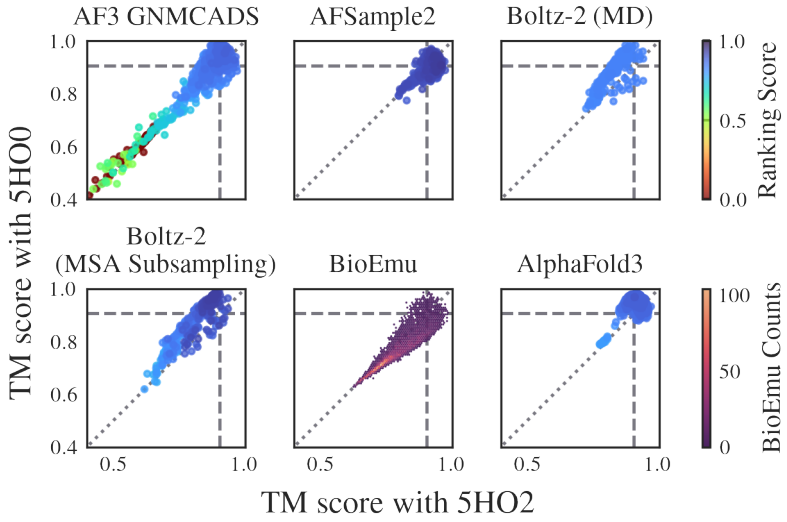
(c) *Left*: Diversity plots of A2RJ53. *Right*: Structures 3FTO (blue), and 3DRF (red).

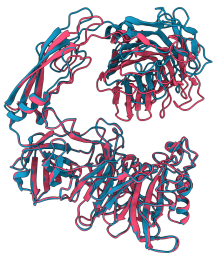
Figure 4: Diversity plots for the Open-Closed dataset targets (continued on next page).



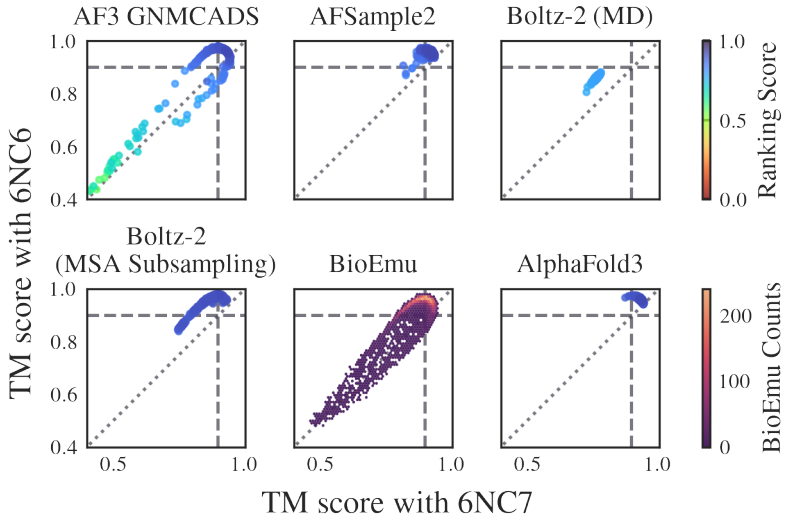


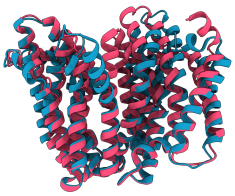
(d) *Left*: Diversity plots of A6UVT1. *Right*: Multimer structures 6HAC (blue), and 6HAE (red).





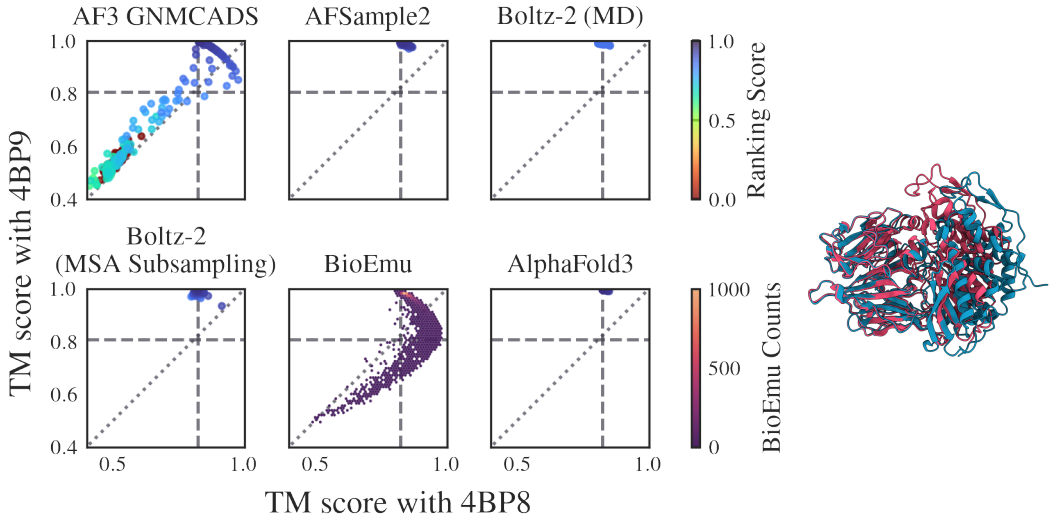
(e) *Left*: Diversity plots of B3EYN2. *Right*: Structures 5HO2 (blue), and 5HO0 (red).



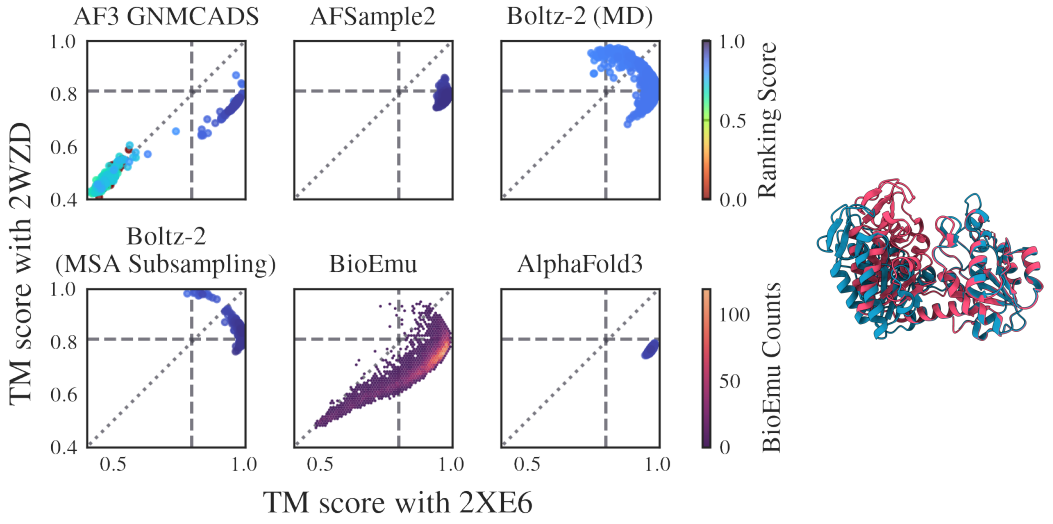


(f) *Left*: Diversity plots of B7IE18. *Right*: Structures 6NC7 (blue), and 6NC6 (red).

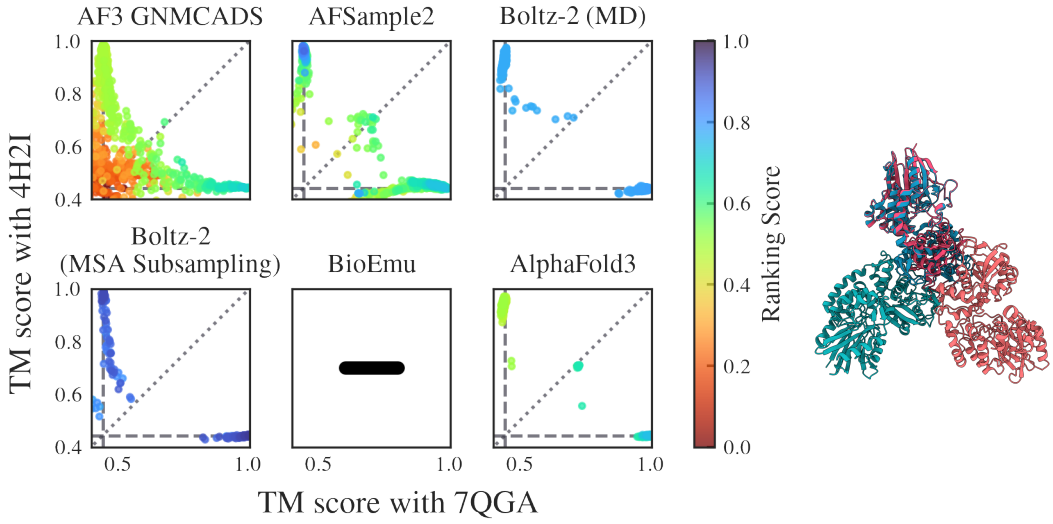
Figure 4: Diversity plots for the Open-Closed dataset targets (continued on next page).



(g) *Left*: Diversity plots of O76728. *Right*: Structures 4BP8 (blue), and 4BP9 (red).

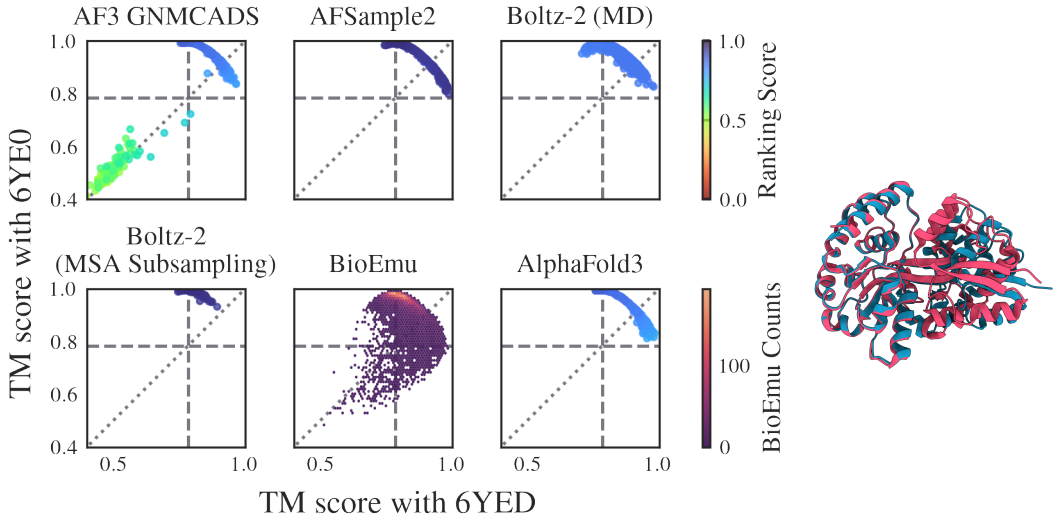


(h) *Left*: Diversity plots of P00558. *Right*: Structures 2XE6 (blue), and 2WZD (red).

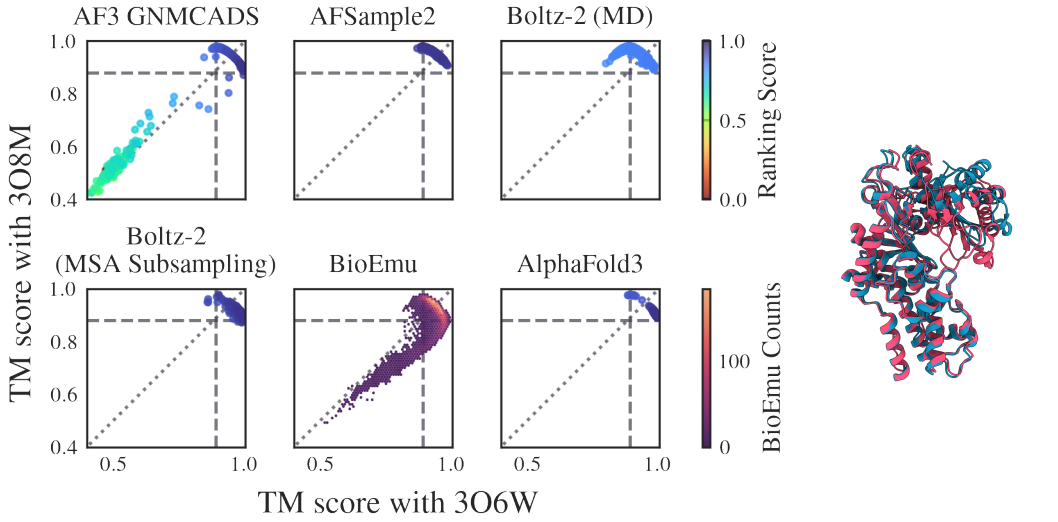


(i) *Left*: Diversity plots of P21589. *Right*: Multimer structures 7QGA (blue), and 4H2I (red).

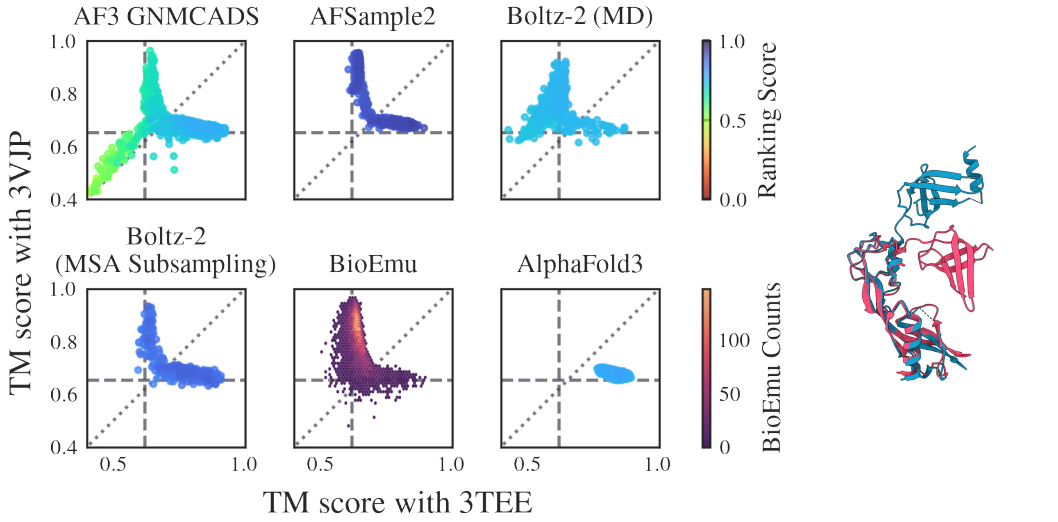
Figure 4: Diversity plots for the Open-Closed dataset targets (continued on next page).



(j) *Left*: Diversity plots of P31133. *Right*: Structures 6YED (blue), and 6YE0 (red).

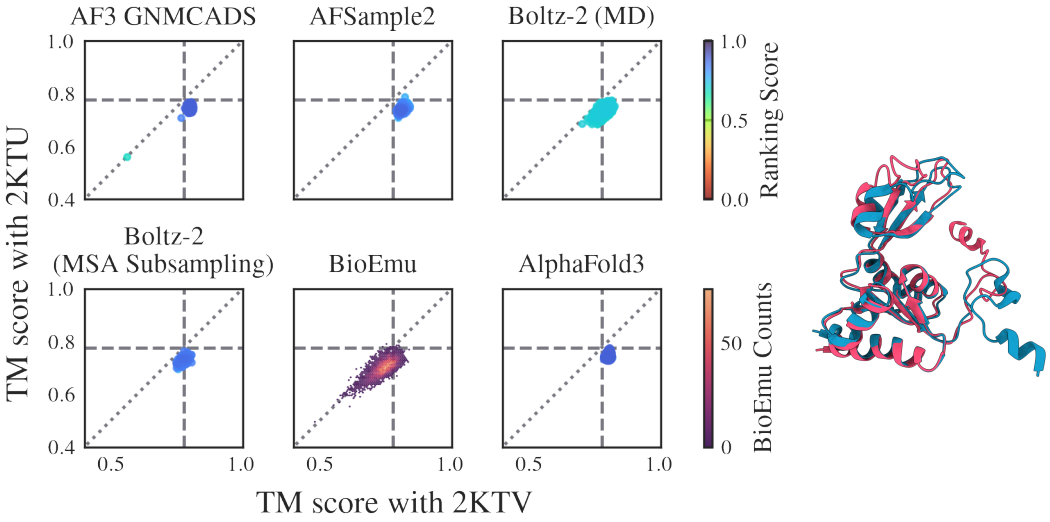


(k) *Left*: Diversity plots of P33284. *Right*: Structures 3O6W (blue), and 3O8M (red).

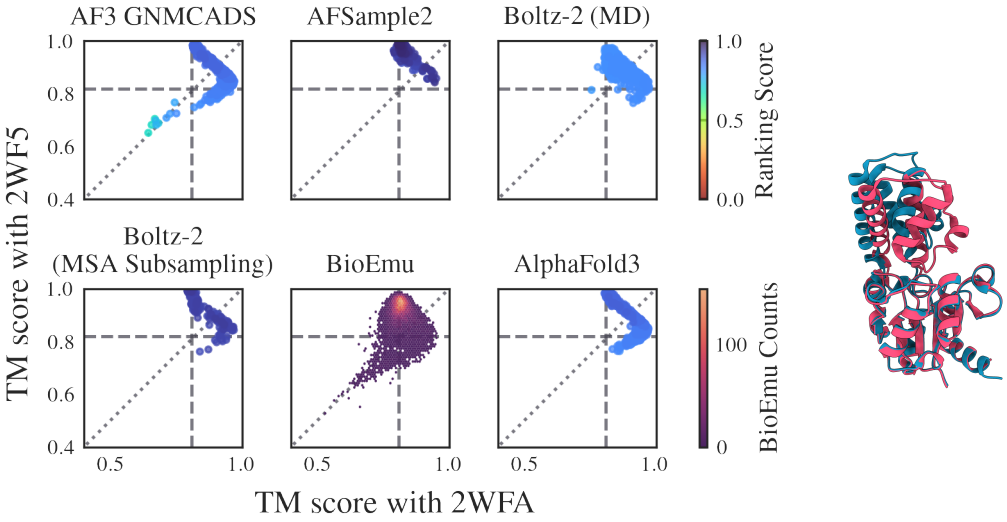


(l) *Left*: Diversity plots of P40131. *Right*: Structures 3TEE (blue), and 3VJP (red).

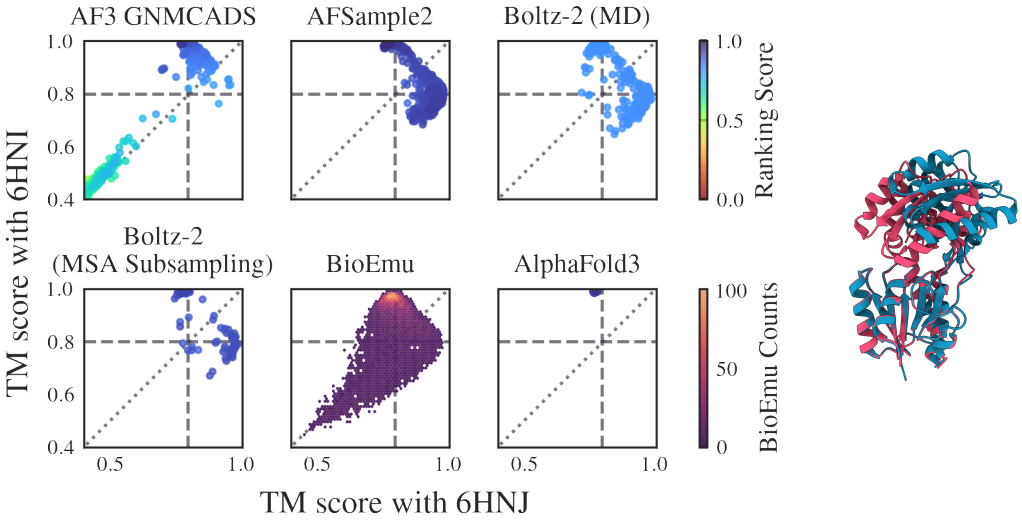
Figure 4: Diversity plots for the Open-Closed dataset targets (continued on next page).



(m) *Left*: Diversity plots of P62495. *Right*: Structures 2KTV (blue), and 2KTU (red).

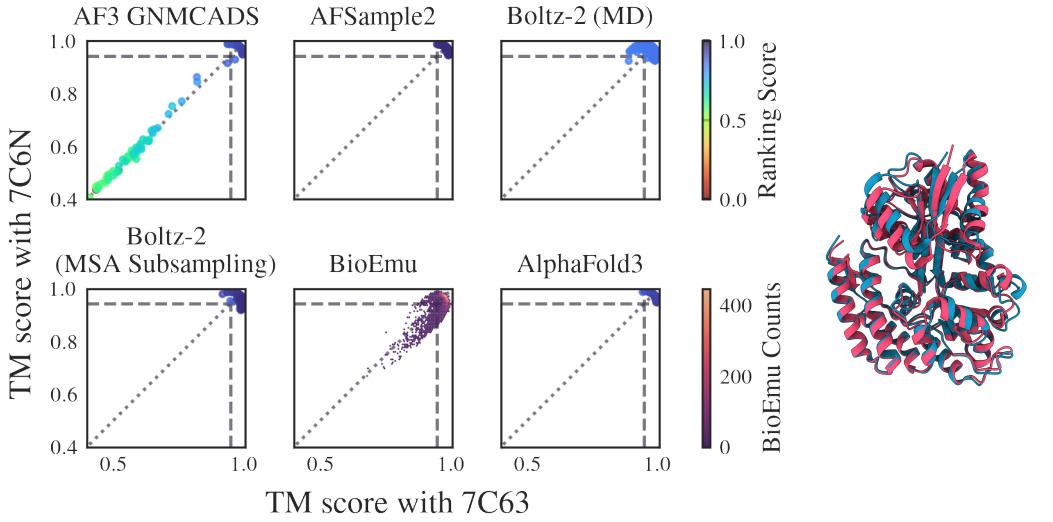


(n) *Left*: Diversity plots of P71447. *Right*: Structures 2WFA (blue), and 2WF5 (red).

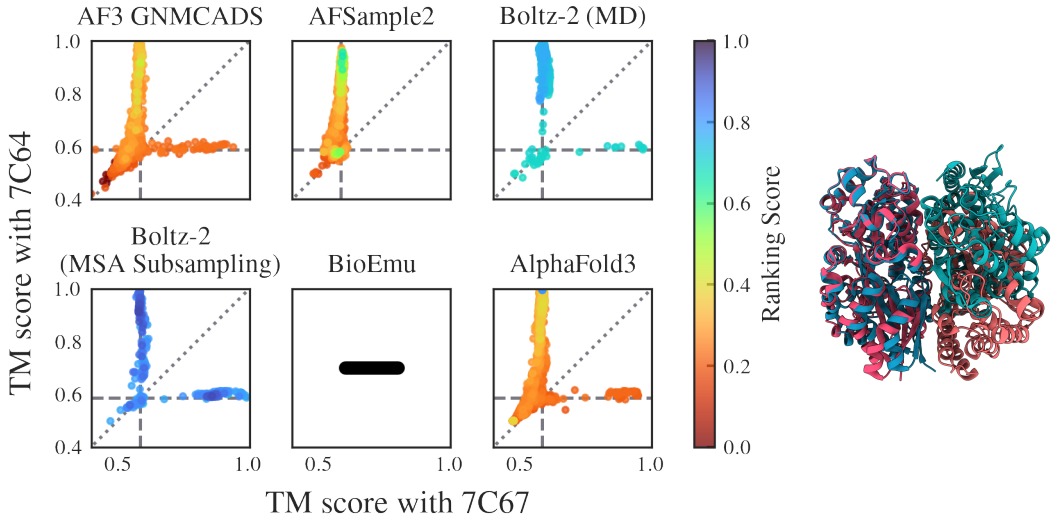


(o) *Left*: Diversity plots of Q18A65. *Right*: Structures 6HNJ (blue), and 6HNI (red).

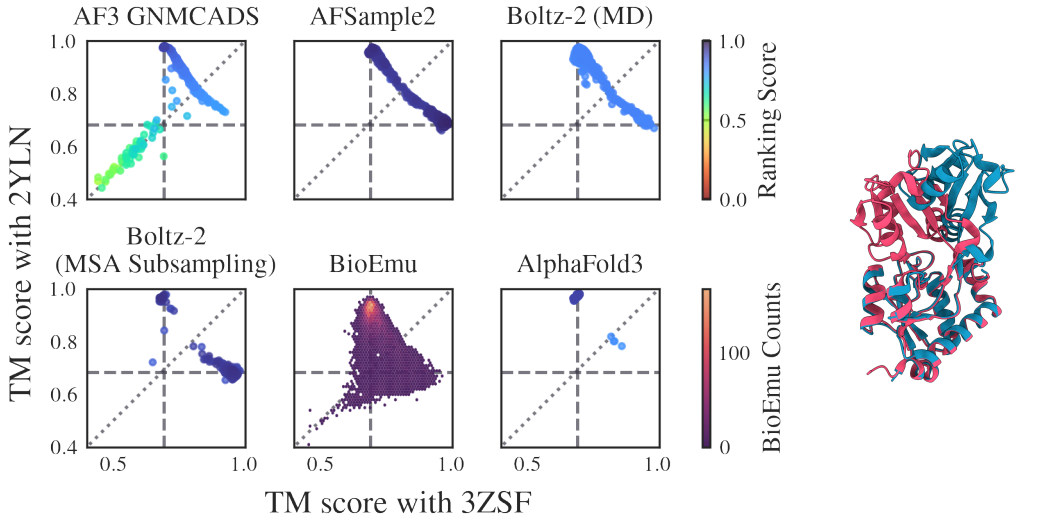
Figure 4: Diversity plots for the Open-Closed dataset targets (continued on next page).



(p) *Left*: Diversity plots of Q53W80. *Right*: Structures 7C63 (blue), and 7C6N (red).

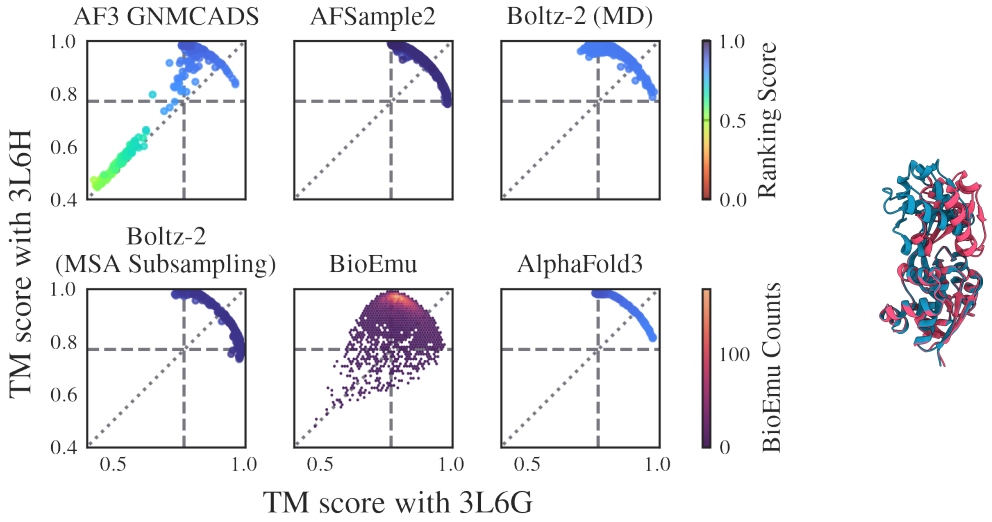


(q) *Left*: Diversity plots of Q53W80. *Right*: Multimer structures 7C67 (blue), and 7C64 (red).

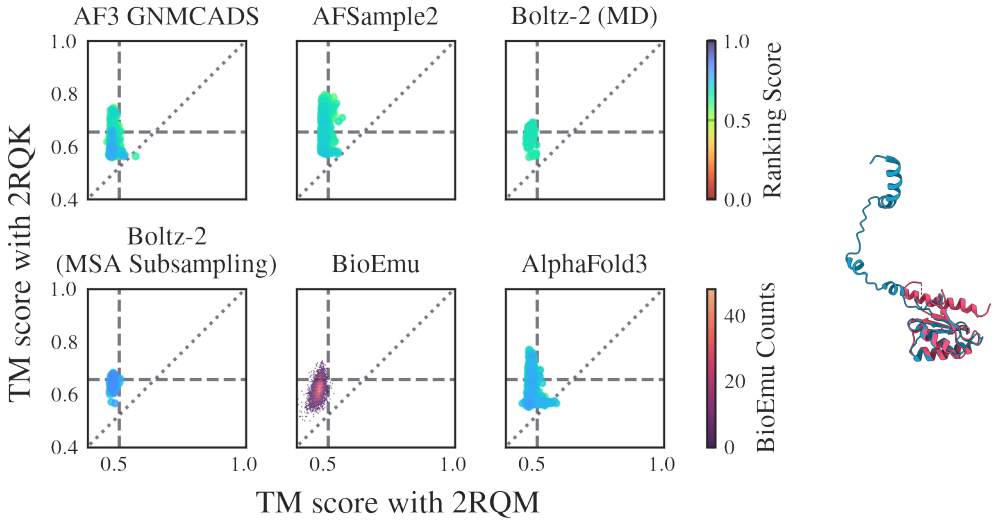


(r) *Left*: Diversity plots of Q5F9M1. *Right*: Structures 3ZSF (blue), and 2YLN (red).

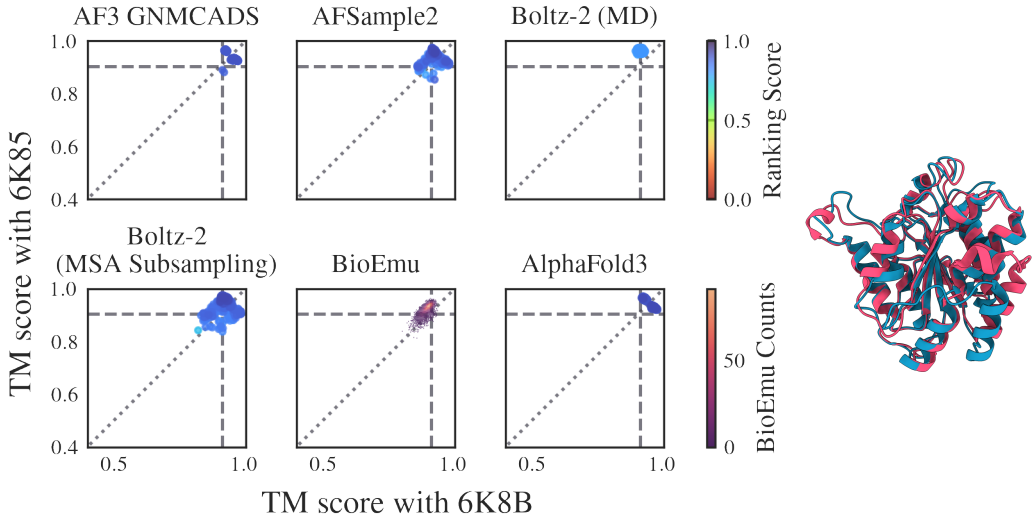
Figure 4: Diversity plots for the Open-Closed dataset targets (continued on next page).



(s) *Left*: Diversity plots of Q7DAU8. *Right*: Structures 3L6G (blue), and 3L6H (red).

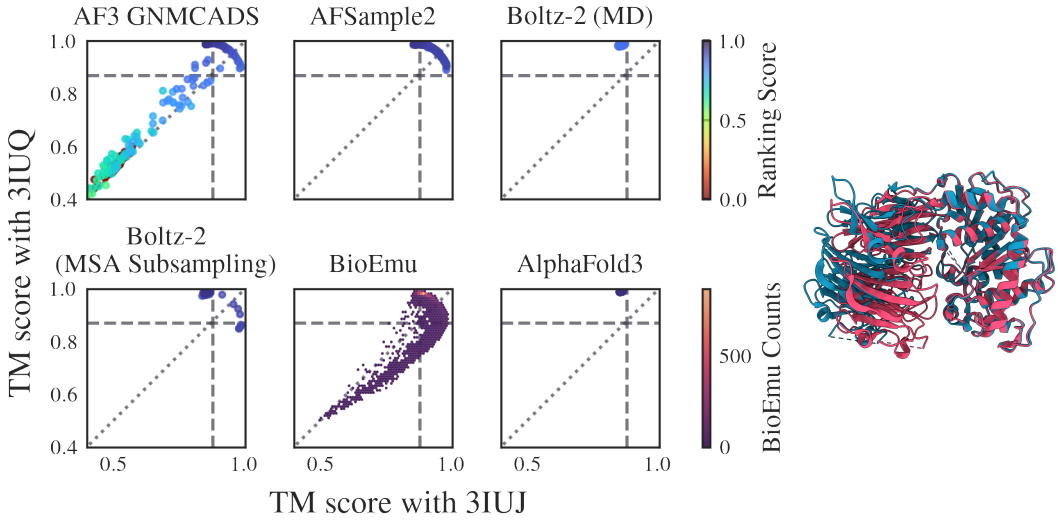


(t) *Left*: Diversity plots of Q9ERE7. *Right*: Structures 2RQM (blue), and 2RQK (red).

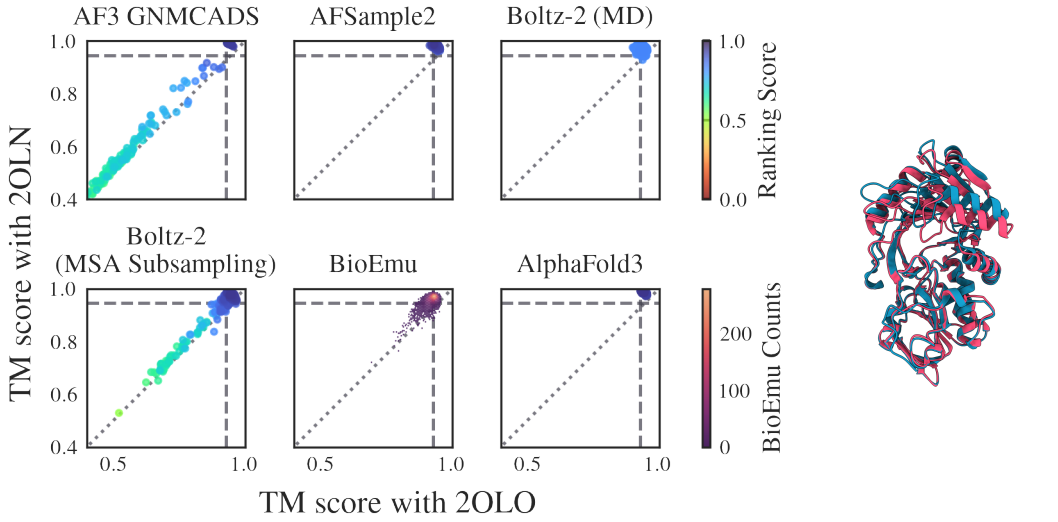


(u) *Left*: Diversity plots of Q9SS90. *Right*: Structures 6K8B (blue), and 6K85 (red).

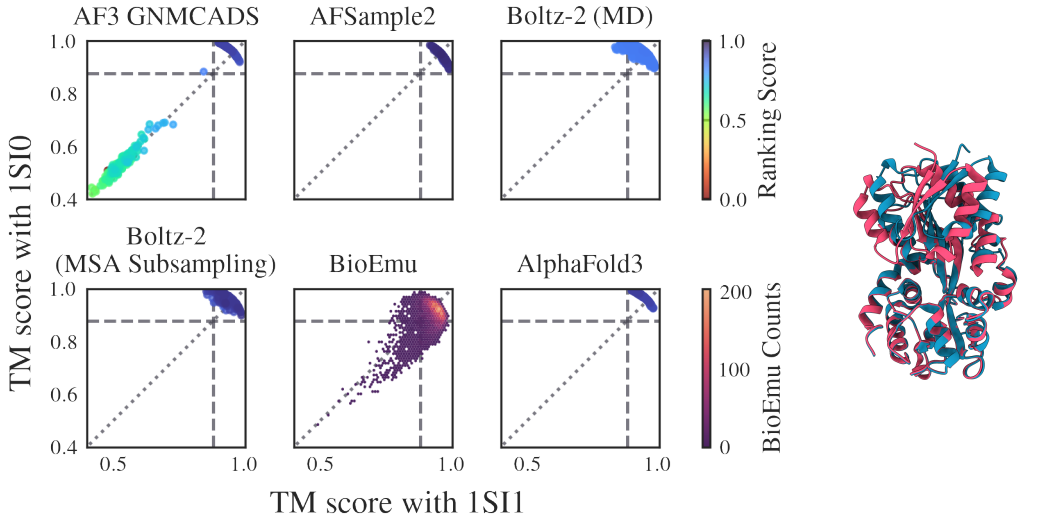
Figure 4: Diversity plots for the Open-Closed dataset targets (continued on next page).



(v) *Left*: Diversity plots of Q9X6R4. *Right*: Structures 3IUJ (blue), and 3IUQ (red).



(w) *Left*: Diversity plots of Q9X9P9. *Right*: Structures 2OLO (blue), and 2OLN (red).



(x) *Left*: Diversity plots of Q9Z4N6. *Right*: Structures 1SI1 (blue), and 1SI0 (red).

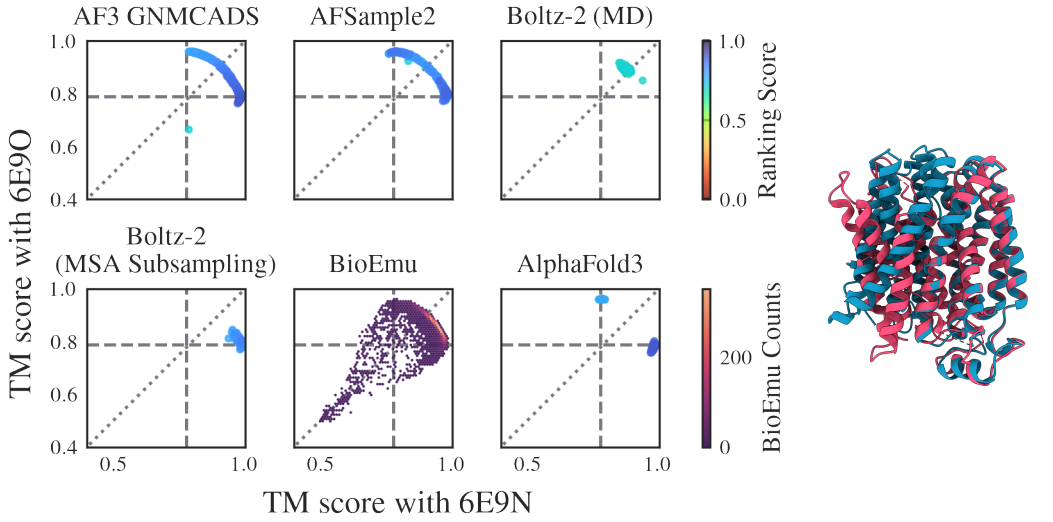
Figure 4: Diversity plots for the Open-Closed dataset targets.

C.2 TRANSPORTERS DATASET

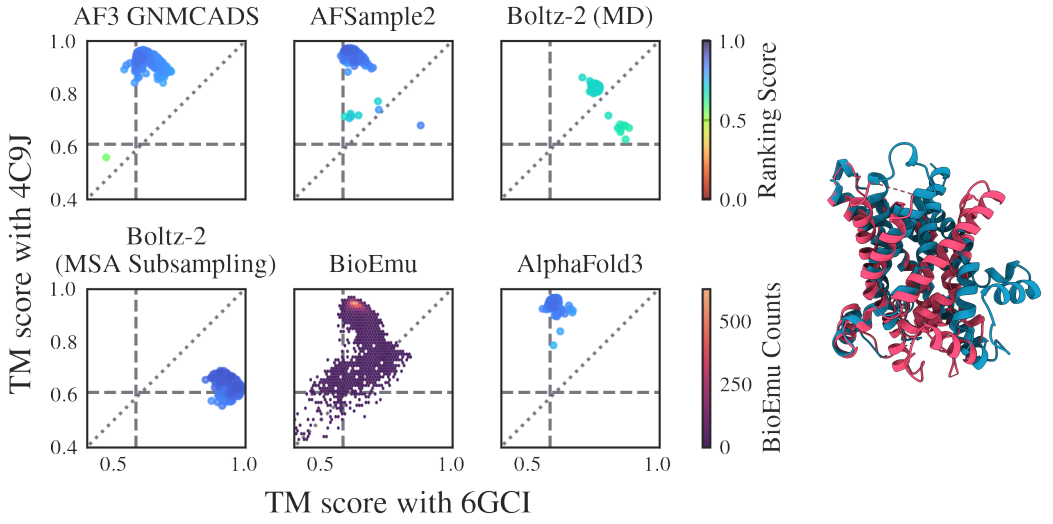
The Transporters dataset contains 15 protein targets with 18 different conformational pairs . In contrast to the version proposed by Xie & Huang (2024) where the transporters that exist in a oligomeric state was predicted by one of its chains, predictions were performed with all of the protein chains of interest in the first biological assembly for each target. The sugar-low affinity conformation of melB with PDB identifier 8T60, the inward open and closed conformations of murJ with PDB identifiers 6NC7 and 6NC6 is appended to the dataset. Similarly, the intermediate outward facing state of SLC1A1 with PDB identifier 8CTC is also analyzed. Additionally, the outward-facing conformation of the target PF0708 is replaced with 6GWH as the structure of 6GWH has been solved by the same study that solved the structure of the inward-facing conformation, Zakrzewska et al. (2019). The details of the targets and the conformations used can be seen in Table 2.

Table 2: Transporters dataset targets

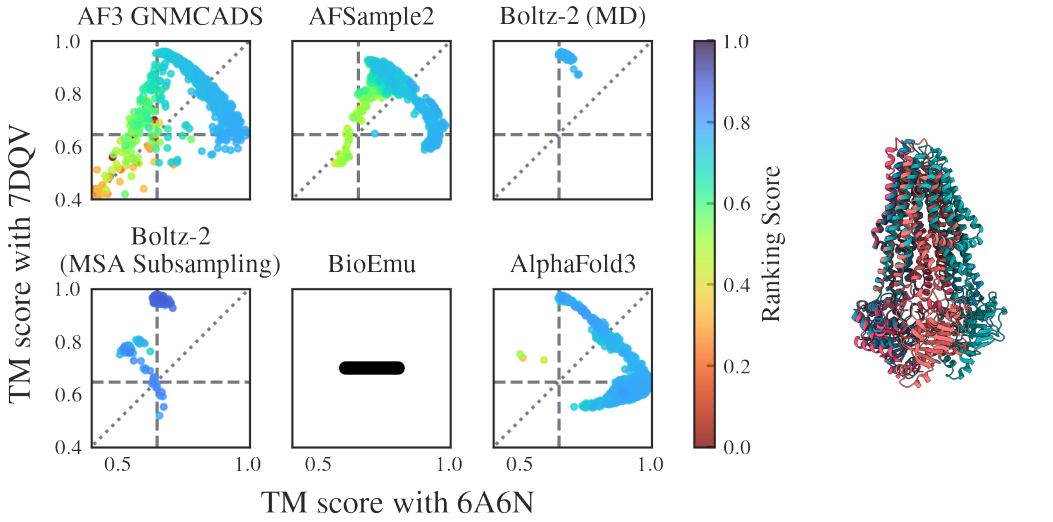
Target Name	Conformation 1 PDB ID	Conformation 2 PDB ID
A5U30_003247	6E9N	6E9O
AAC3	6GCI	4C9J
CYME_CMD148C	6A6N	7DQV
MFSD2A	7N98	7MJS
PF0708	6FHZ	6GWH
SLC1A1	6S3Q	6X3E
SLC1A1	6S3Q	8CTC
SPF1	6XMS	6XMU
TM_0287	3QF4	6QV1
TT_C0976	5MKK	6RAJ
mdfA	6VS1	6GV1
melB	7L17	4M64
melB	7L17	8T60
murJ	6NC7	6NC6
murJ	6NC7	6NC9
ptsG	6BVG	5IWS
slc39	6BTX	5AYM
wlaB	5C78	6HRC



(a) *Left*: Diversity plots of A5U30_003247. *Right*: Structures of 6E9N (blue), and 6E9O (red).

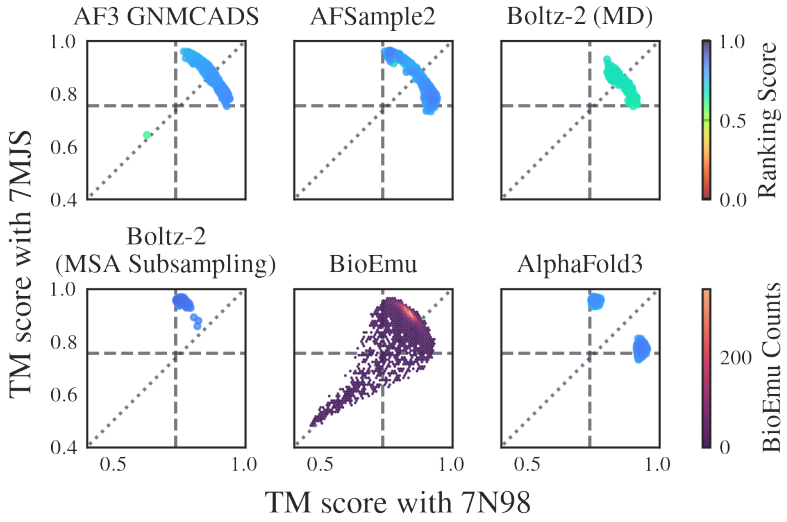


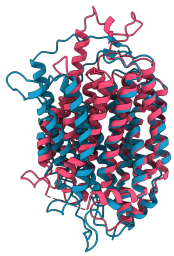
(b) *Left*: Diversity plots of AAC3. *Right*: Structures of 6GCI (blue), and 4C9J (red).



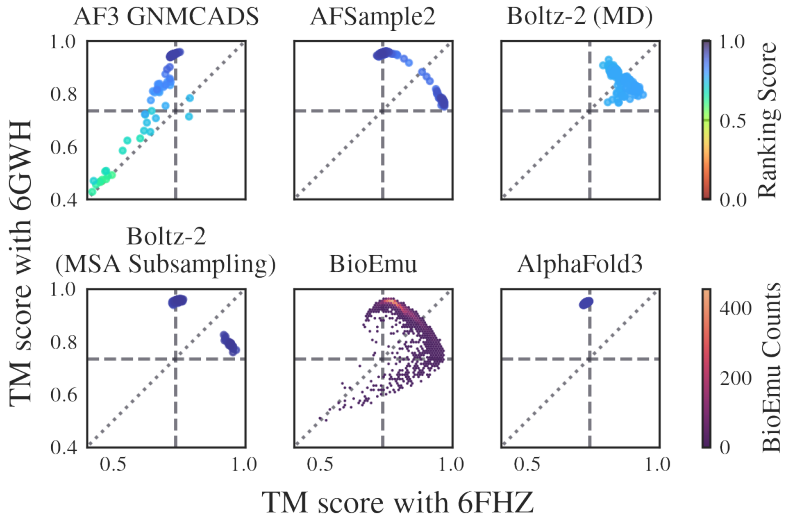
(c) *Left*: Diversity plots of CYME_CMD148C. *Right*: Multimer structures of 6A6N (blue), and 7DQV (red).

Figure 5: Diversity plots for the Transporters dataset targets (continued on next page).

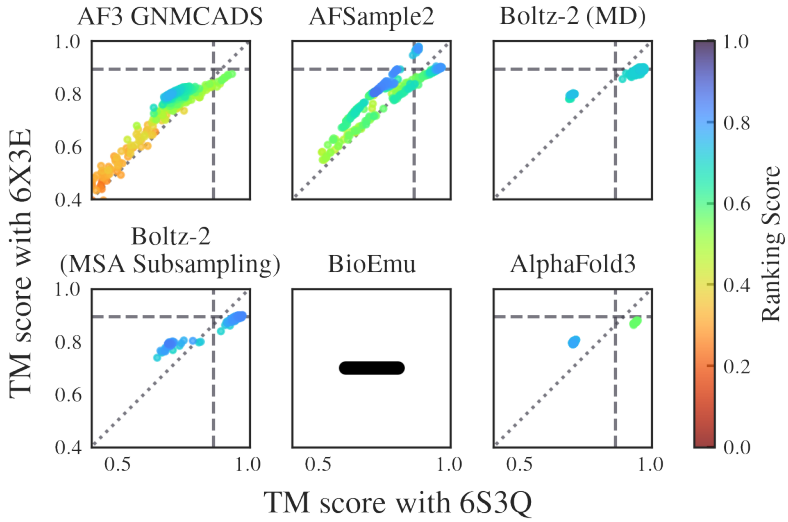


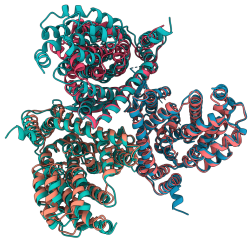


(d) *Left*: Diversity plots of MFSD2A. *Right*: Structures of 7N98 (blue), and 7MJS (red).



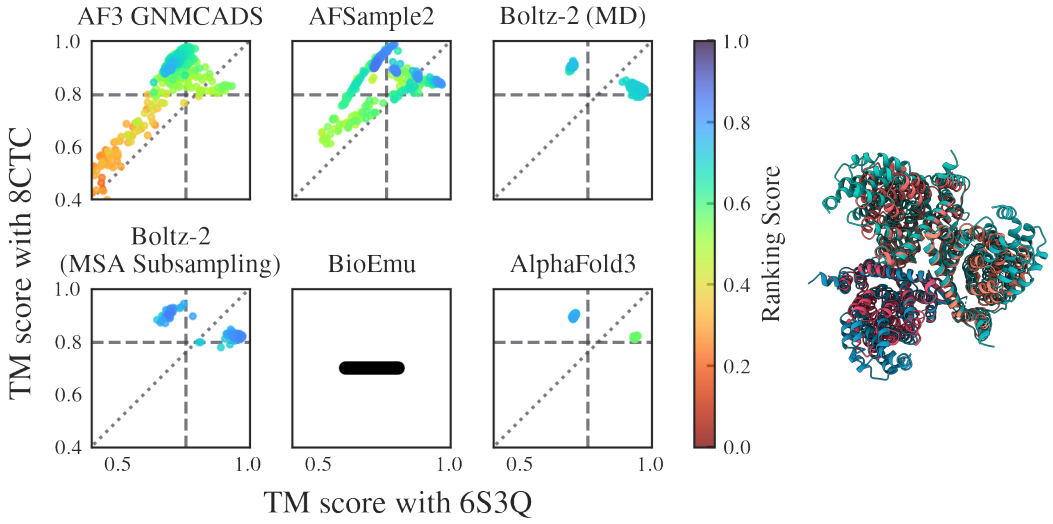
(e) *Left*: Diversity plots of PF0708. *Right*: Structures of 6FHZ (blue), and 6GWH (red).



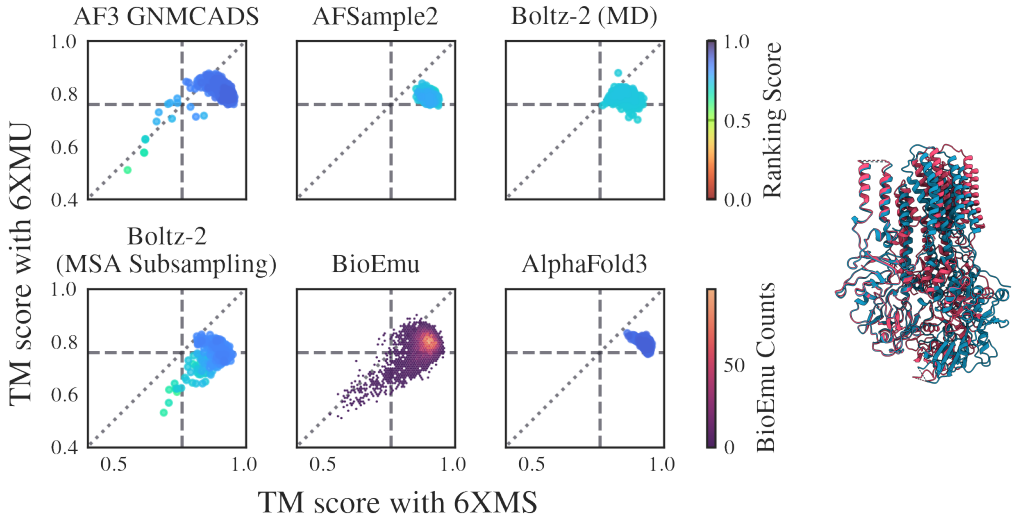


(f) *Left*: Diversity plots of SLC1A1. *Right*: Multimer structures of 6S3Q (blue), and 6X3E (red).

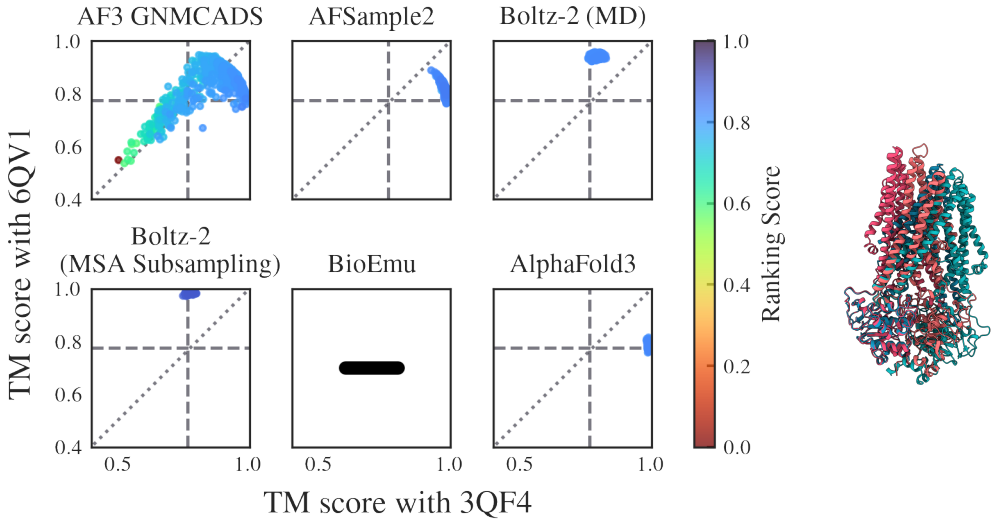
Figure 5: Diversity plots for the Transporters dataset targets (continued on next page).



(g) *Left*: Diversity plots of SLC1A1. *Right*: Multimer structures of 6S3Q (blue), and 8CTC (red).

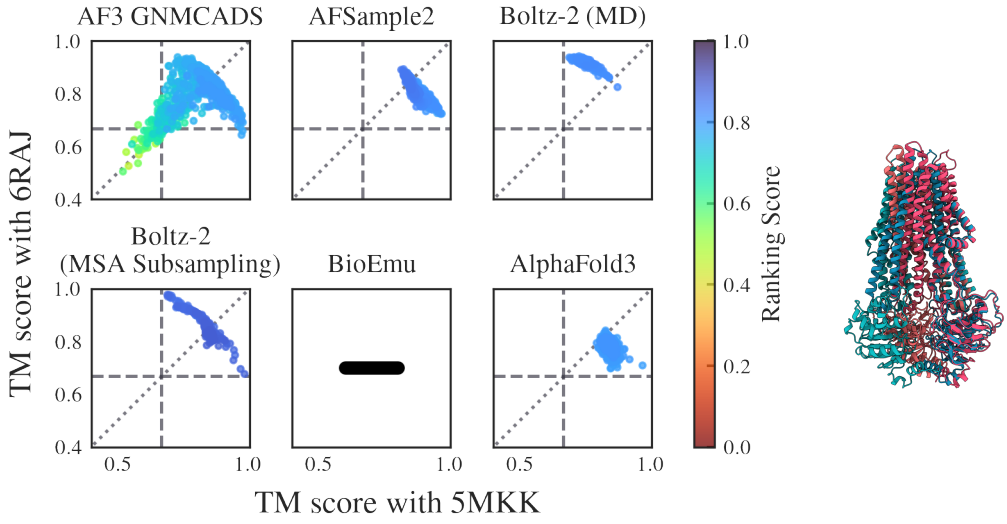


(h) *Left*: Diversity plots of SPF1. *Right*: Structures of 6XMS (blue), and 6XMU (red).

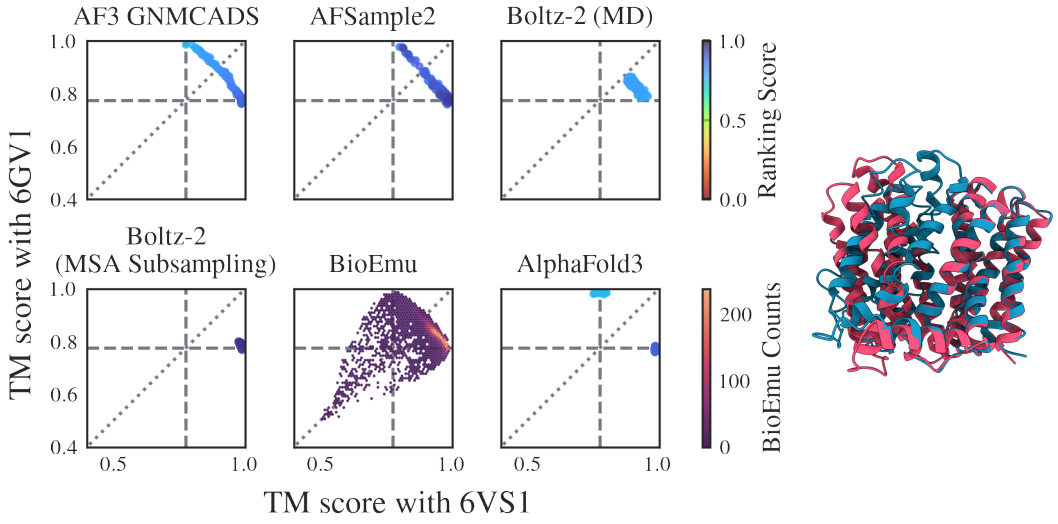


(i) *Left*: Diversity plots of TM_0287. *Right*: Multimer structures of 3QF4 (blue), and 6QV1 (red).

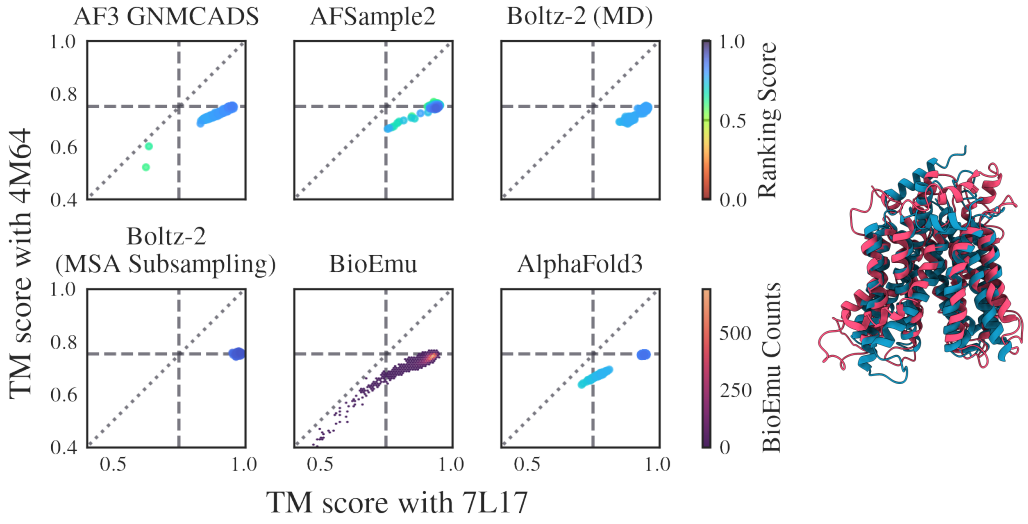
Figure 5: Diversity plots for the Transporters dataset targets (continued on next page).



(j) *Left*: Diversity plots of TT_C0976. *Right*: Multimer structures of 5MKK (blue), and 6RAJ (red).

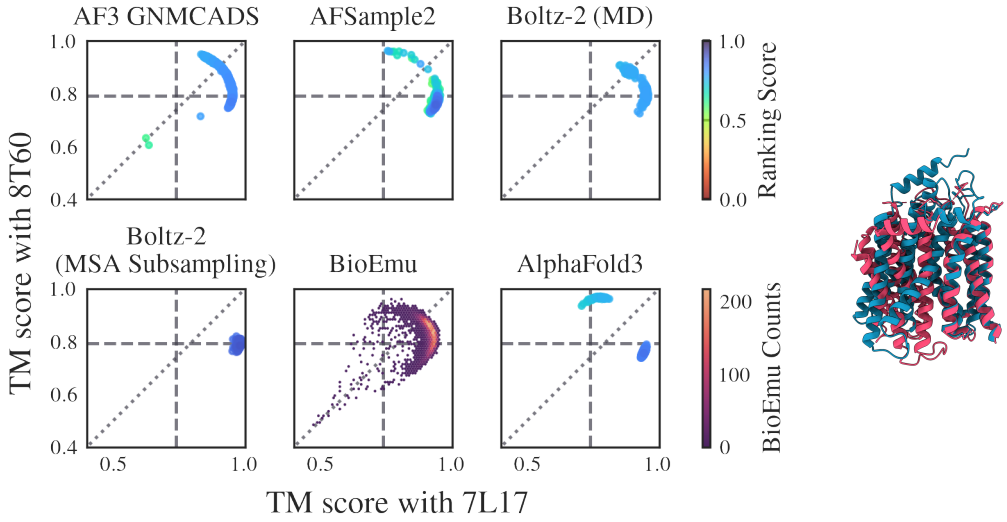


(k) *Left*: Diversity plots of mdfA. *Right*: Structures of 6VS1 (blue), and 6GV1 (red).

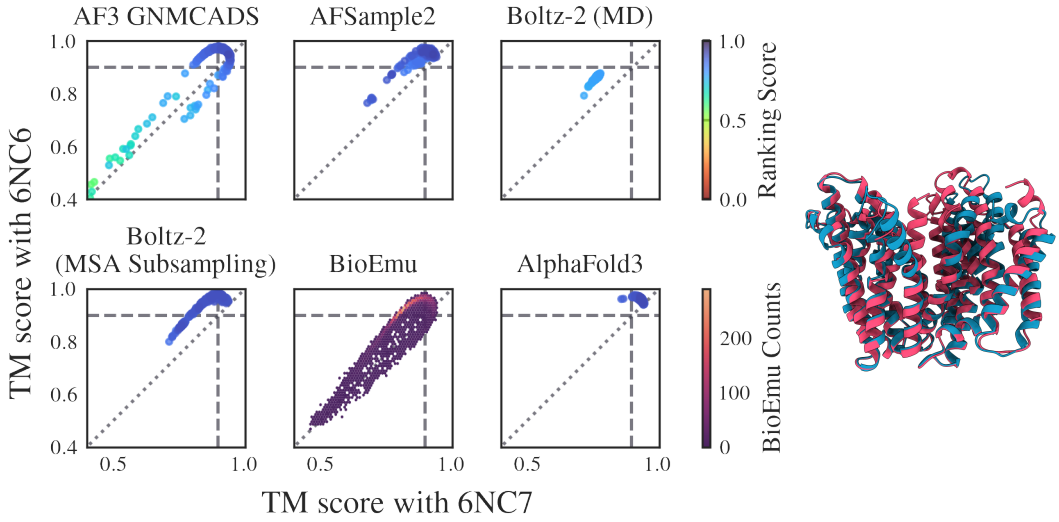


(l) *Left*: Diversity plots of melB. *Right*: Structures of 7L17 (blue), and 4M64 (red).

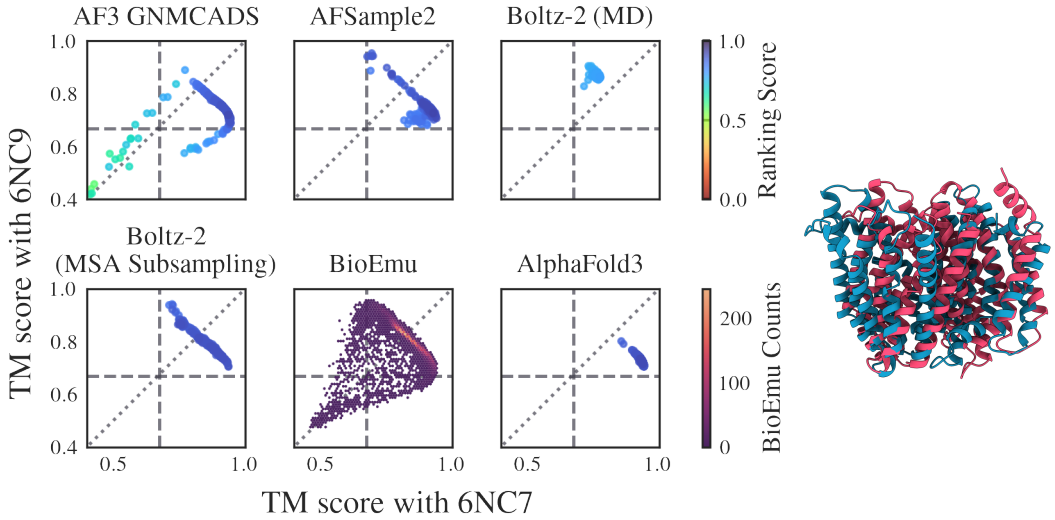
Figure 5: Diversity plots for the Transporters dataset targets (continued on next page).



(m) *Left*: Diversity plots of melB. *Right*: Structures of 7L17 (blue), and 8T60 (red).

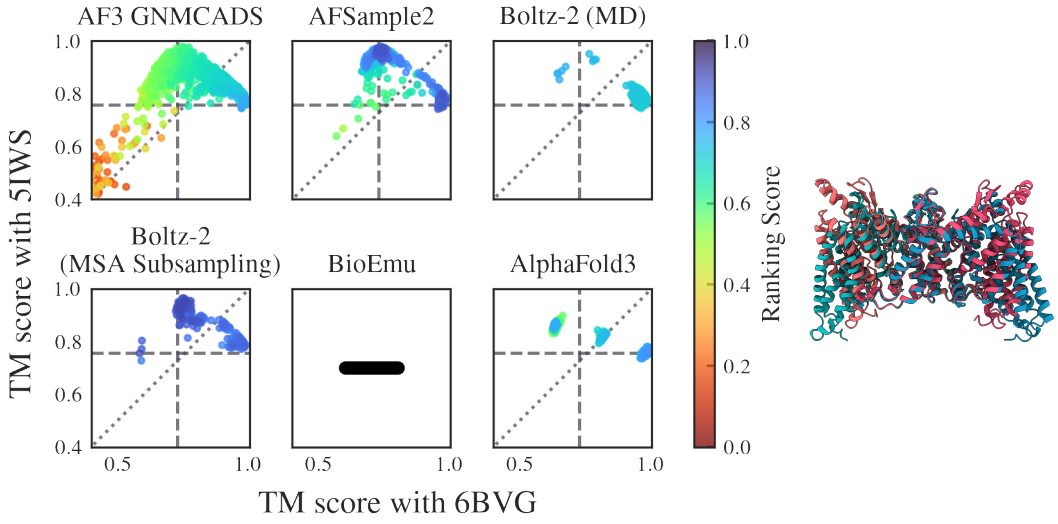


(n) *Left*: Diversity plots of murJ. *Right*: Structures of 6NC7 (blue), and 6NC6 (red).

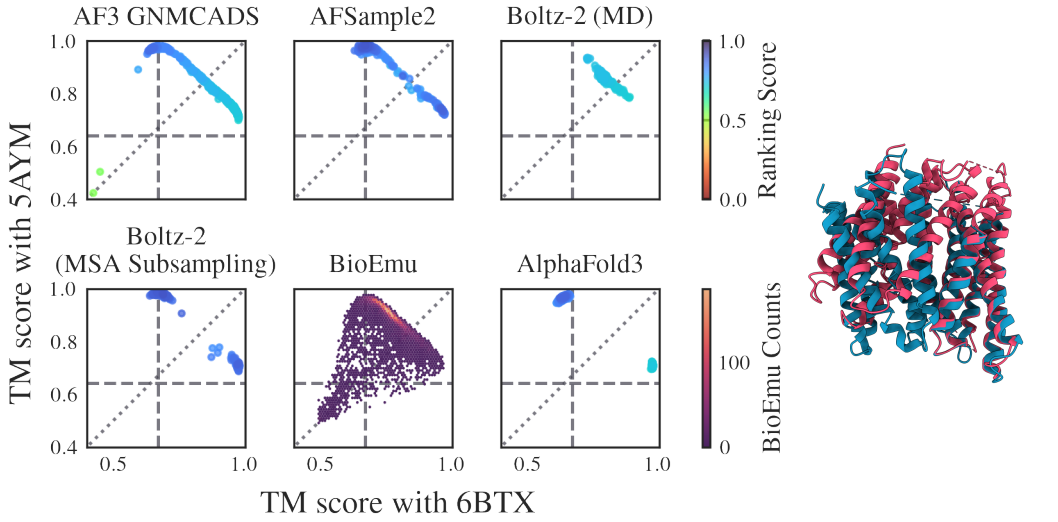


(o) *Left*: Diversity plots of murJ. *Right*: Structures of 6NC7 (blue), and 6NC9 (red).

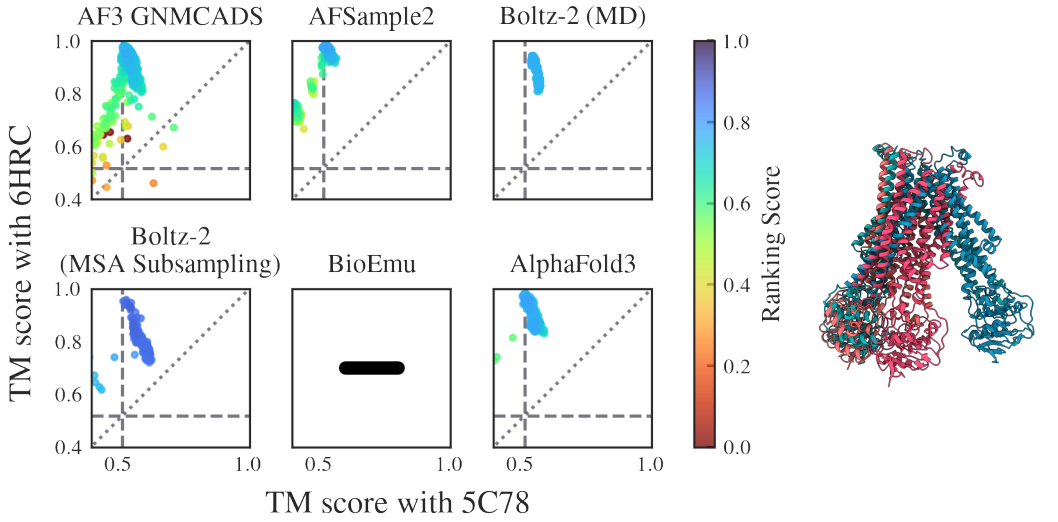
Figure 5: Diversity plots for the Transporters dataset targets (continued on next page).



(p) *Left*: Diversity plots of ptsG. *Right*: Multimer structures of 6BVG (blue), and 5IWS (red).



(q) *Left*: Diversity plots of slc39. *Right*: Structures of 6BTX (blue), and 5AYM (red).



(r) *Left*: Diversity plots of wlaB. *Right*: Multimer structures of 5C78 (blue), and 6HRC (red).

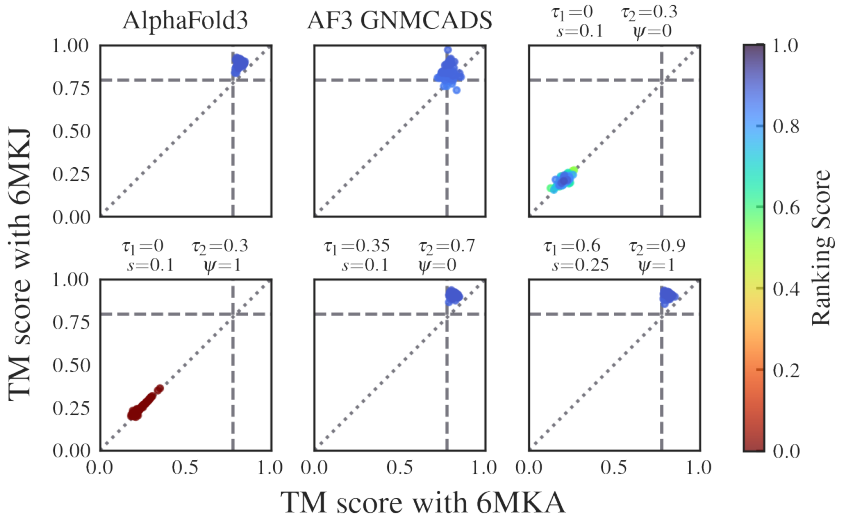
Figure 5: Diversity plots for the Transporters dataset targets.

D ABLATION STUDIES

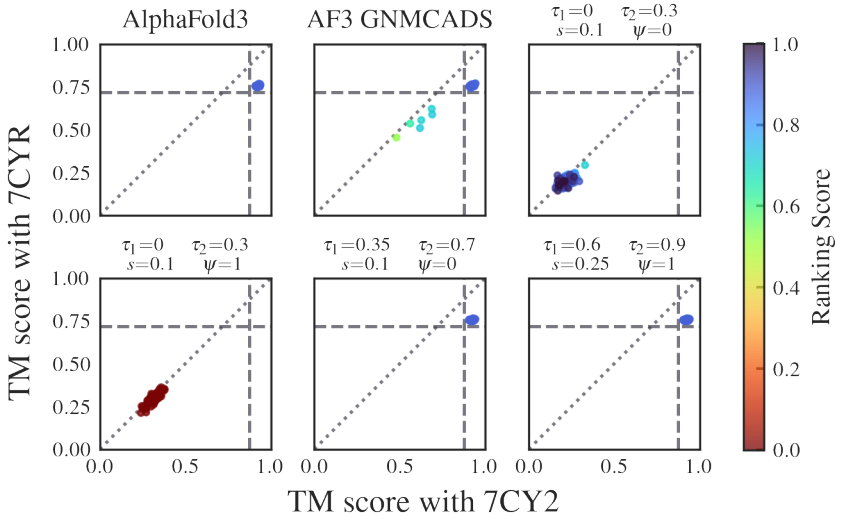
For ablation studies, predictions were performed with 100 samples in contrast to 1000 samples which is used for comparisons with other conformational landscape sampling methods in Appendix C.

D.1 USING CADS WITHOUT GNM

The application of CADS to AlphaFold3's diffusion module directly do not generate correct structures. Although GNMCADS does not use ψ, the mixing factor, it is used in these CADS only runs.

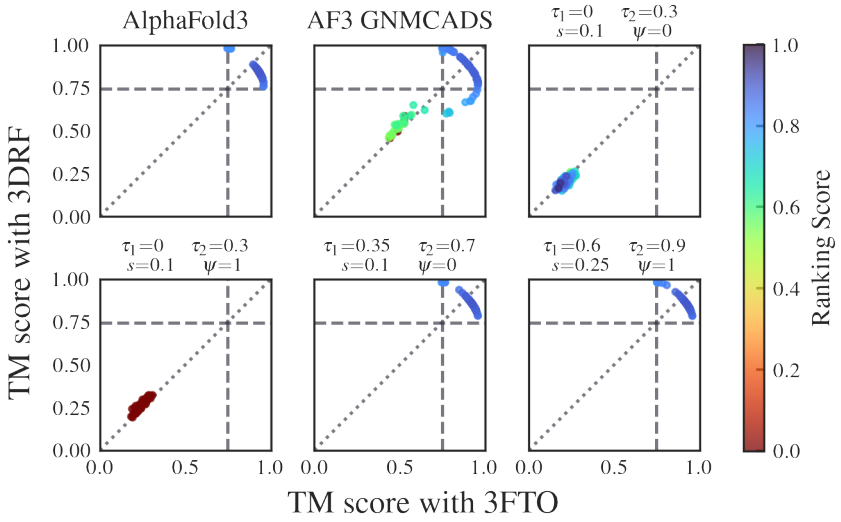


(a) Diversity plots of A0A075Q0W3 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

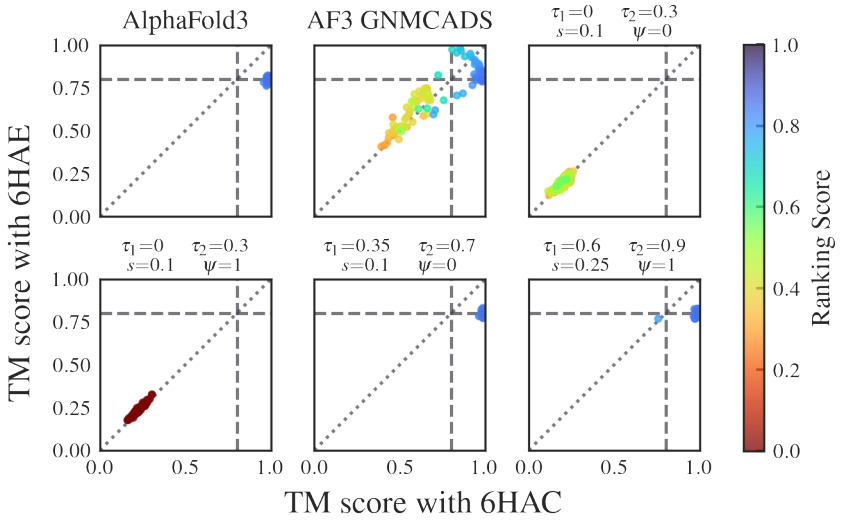


(b) Diversity plots of A0QTT2 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

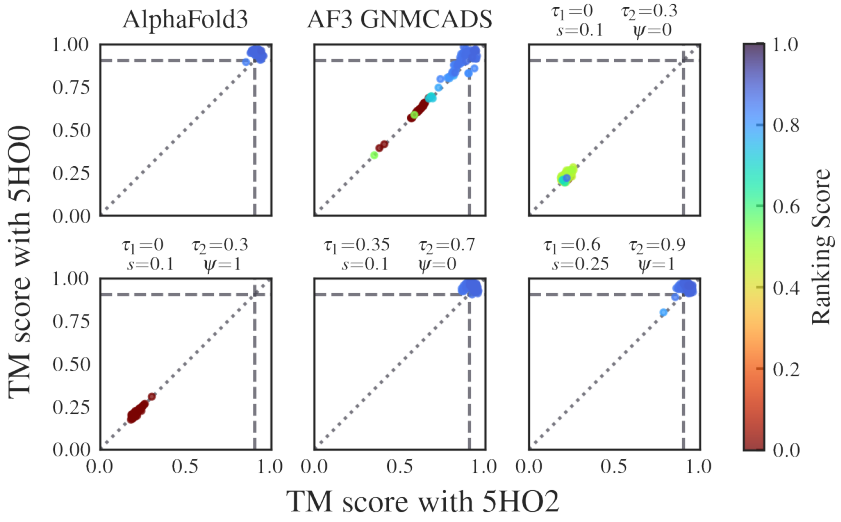
Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



(c) Diversity plots of A2RJ53 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

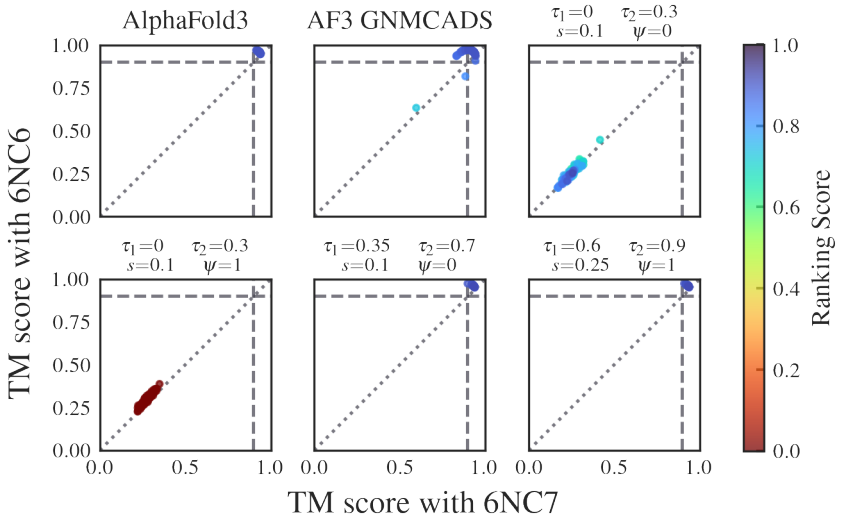


(d) Diversity plots of A6UVT1 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

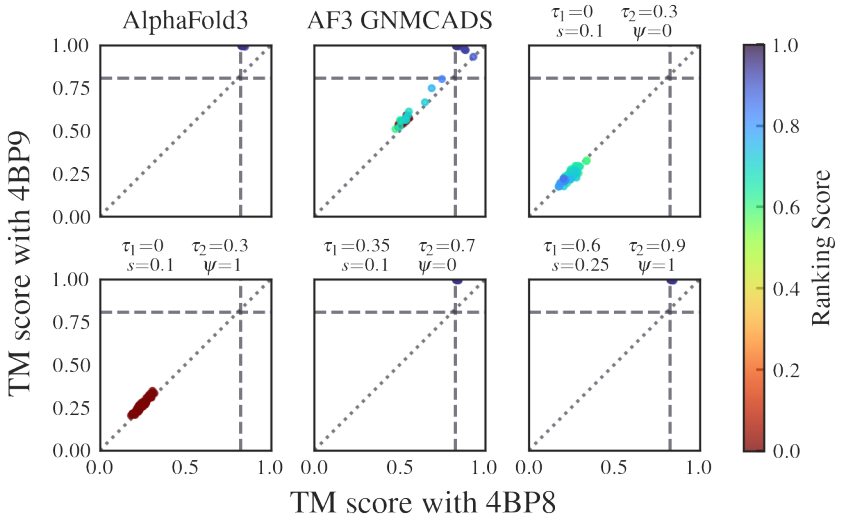


(e) Diversity plots of B3EYN2 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

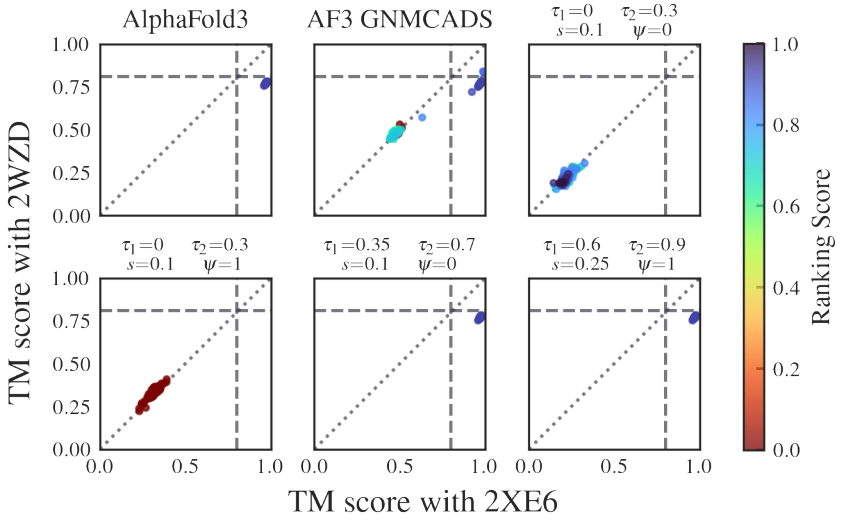
Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



(f) Diversity plots of B7IE18 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

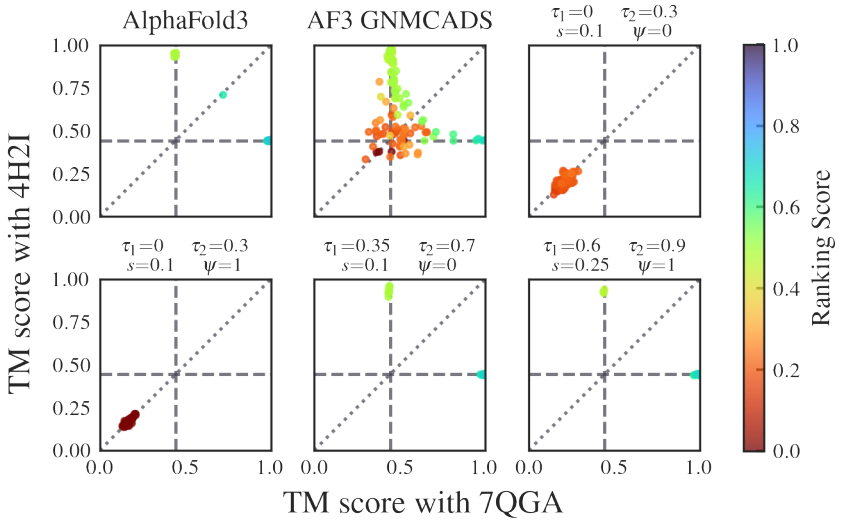


(g) Diversity plots of O76728 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

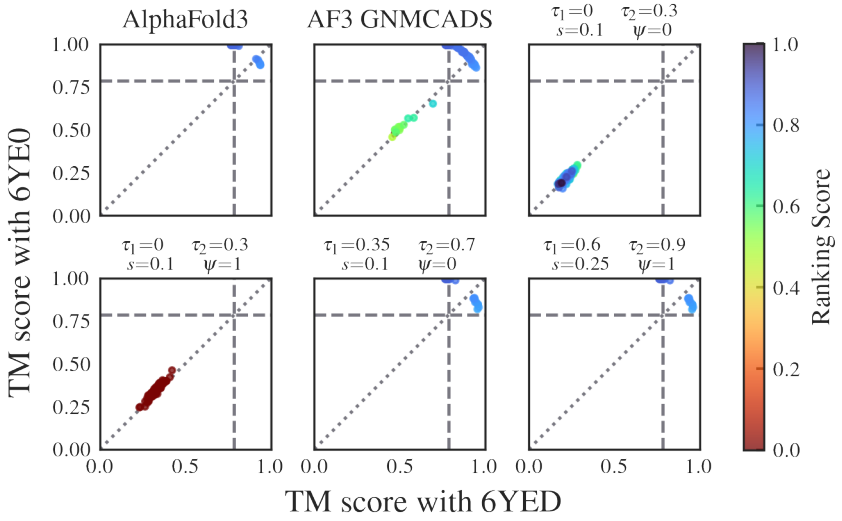


(h) Diversity plots of P00558 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

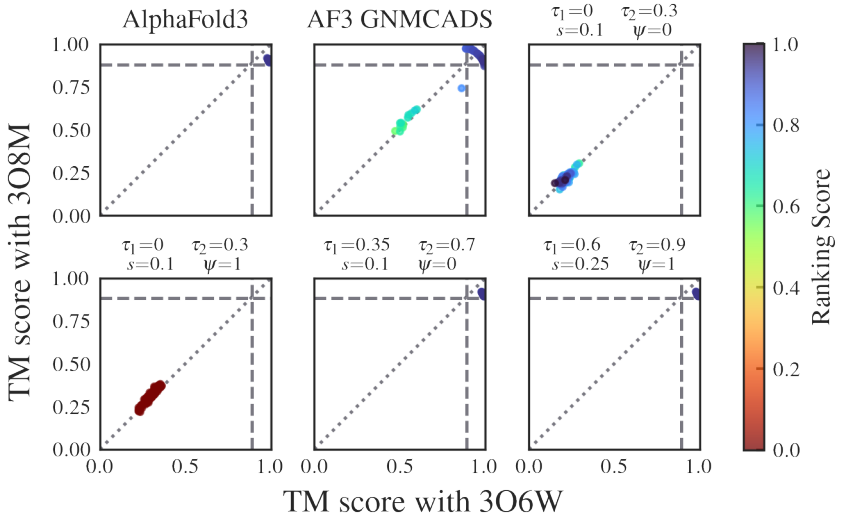
Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



(i) Diversity plots of P21589 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

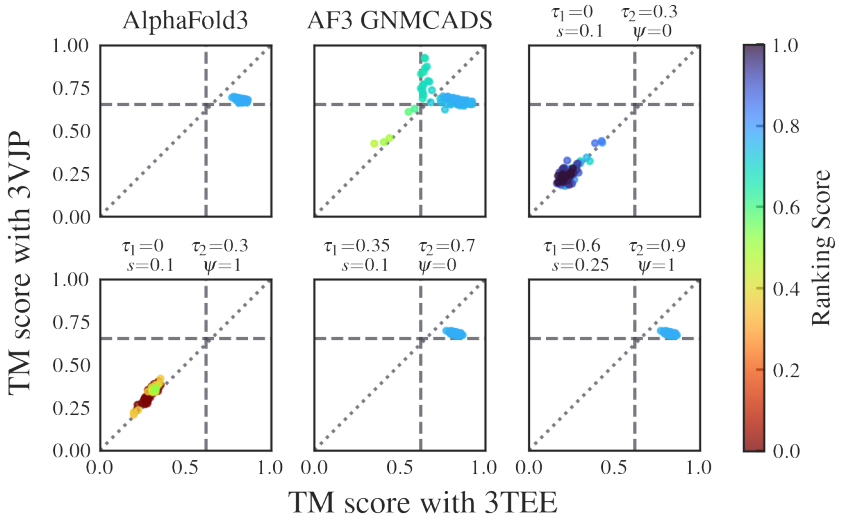


(j) Diversity plots of P31133 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

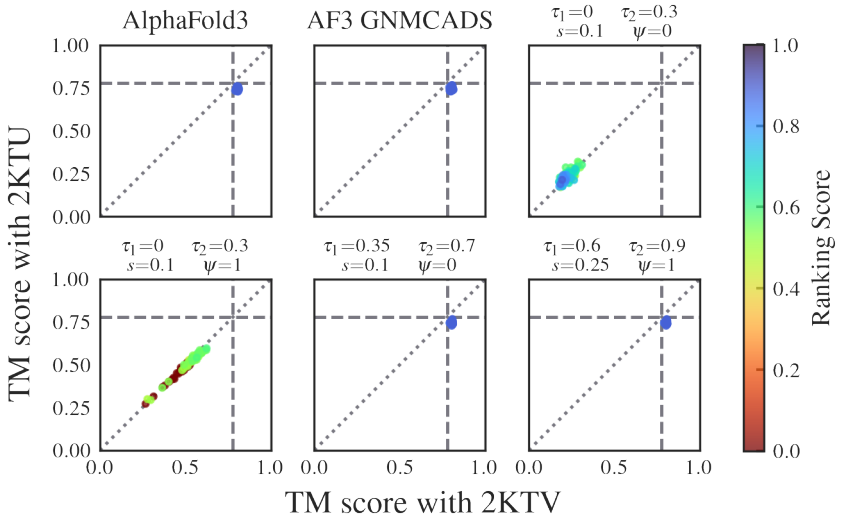


(k) Diversity plots of P33284 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

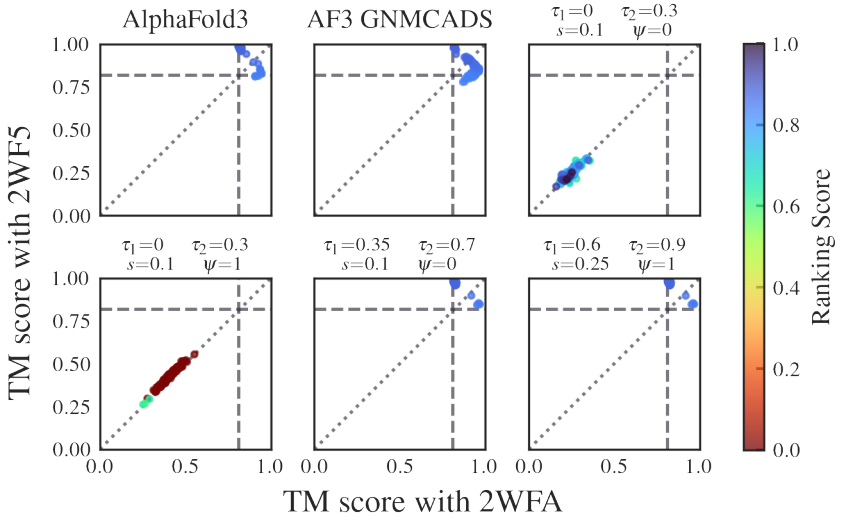
Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



(l) Diversity plots of P40131 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

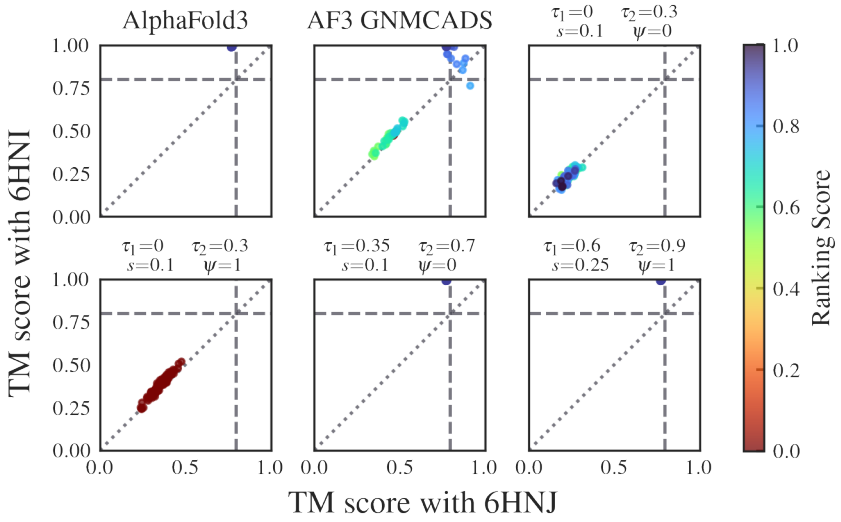


(m) Diversity plots of P62495 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

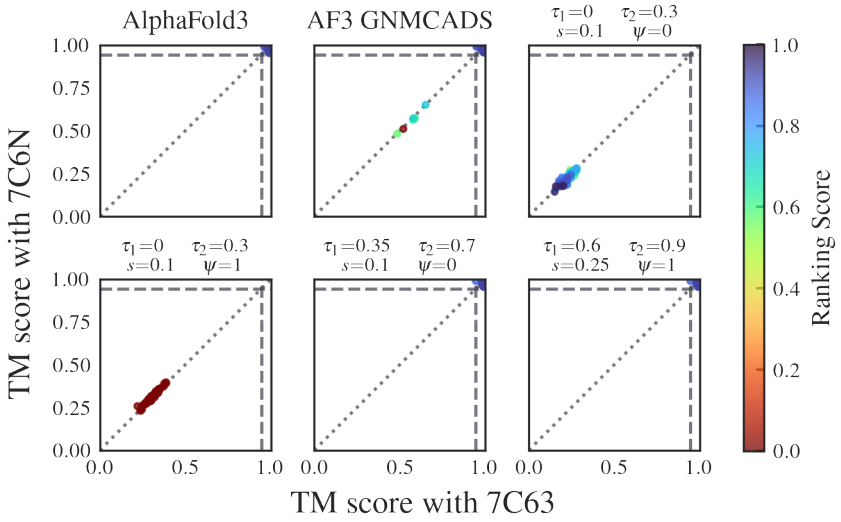


(n) Diversity plots of P71447 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

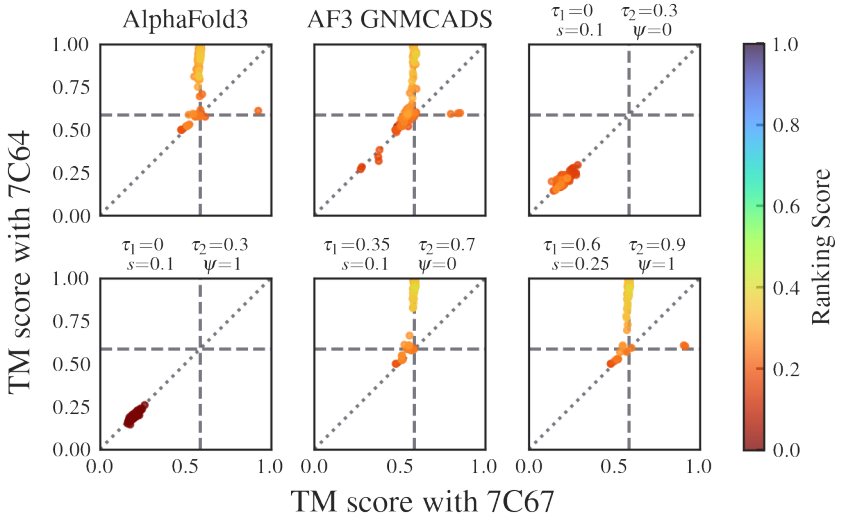
Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



(o) Diversity plots of Q18A65 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

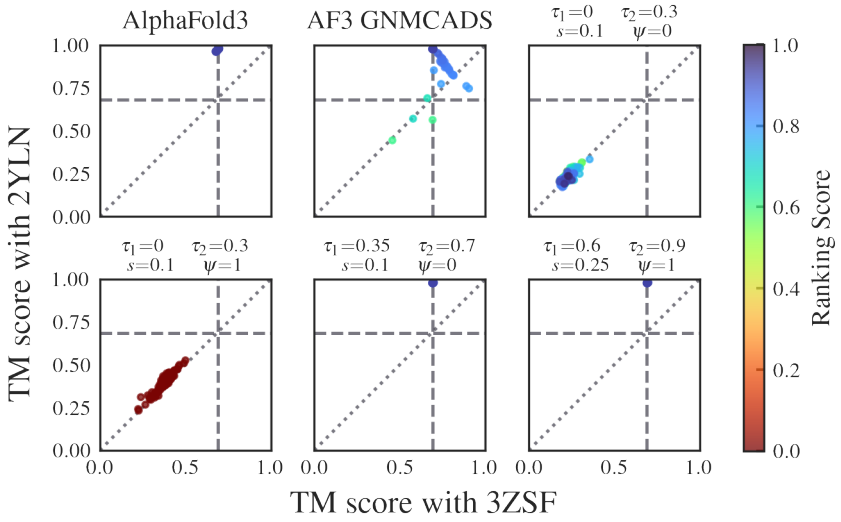


(p) Diversity plots of Q53W80 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

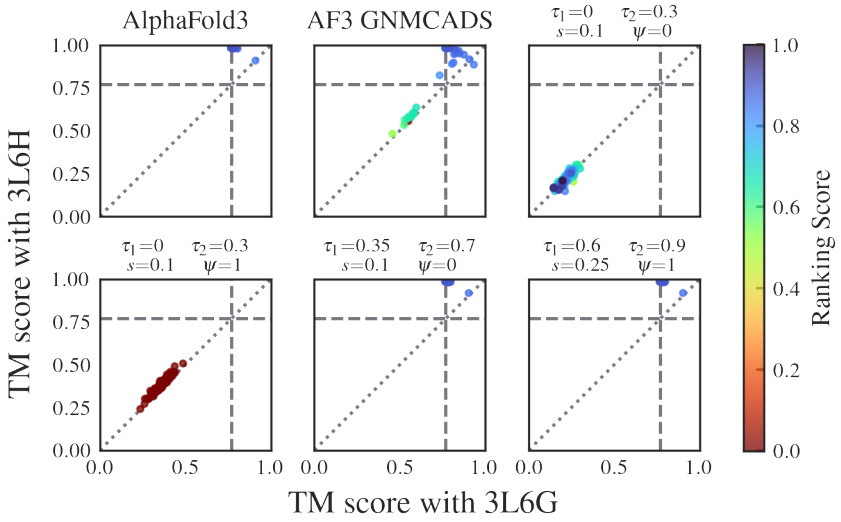


(q) Diversity plots of Q53W80 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

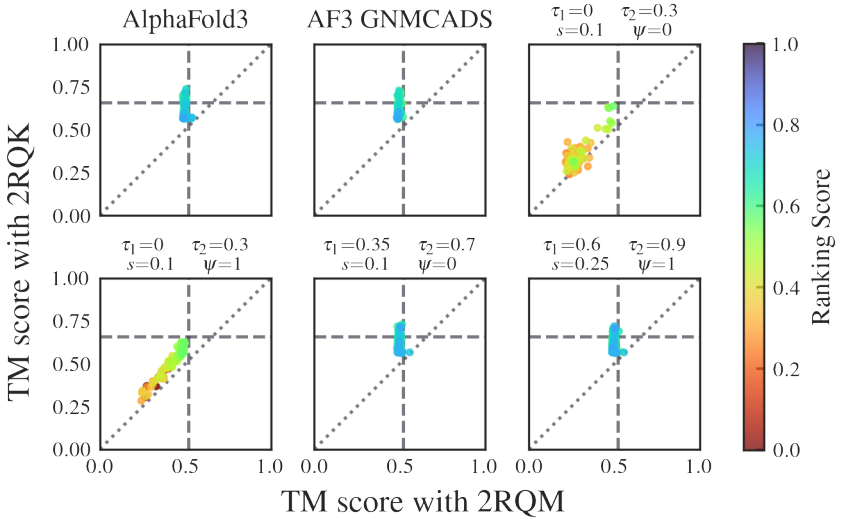
Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



(r) Diversity plots of Q5F9M1 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

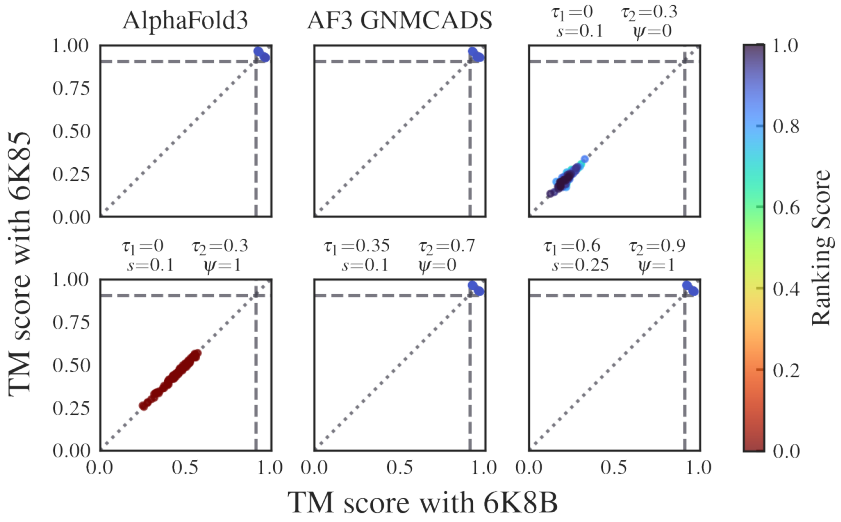


(s) Diversity plots of Q7DAU8 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

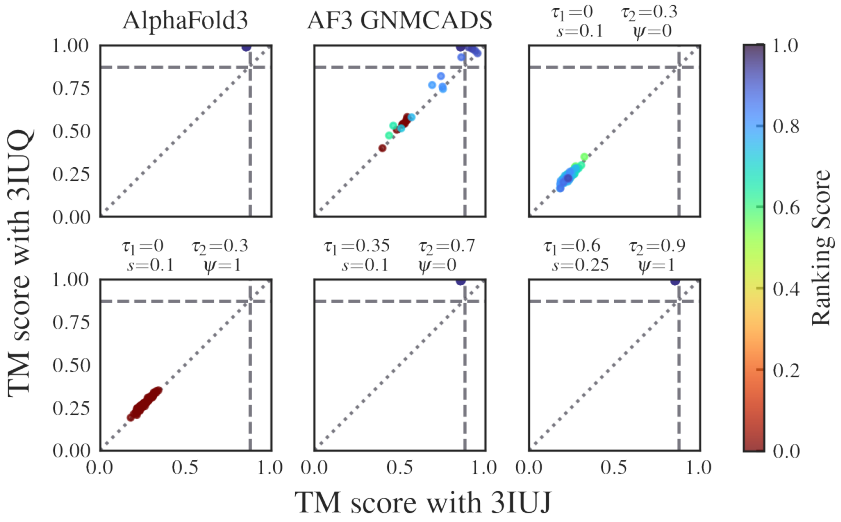


(t) Diversity plots of Q9ERE7 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

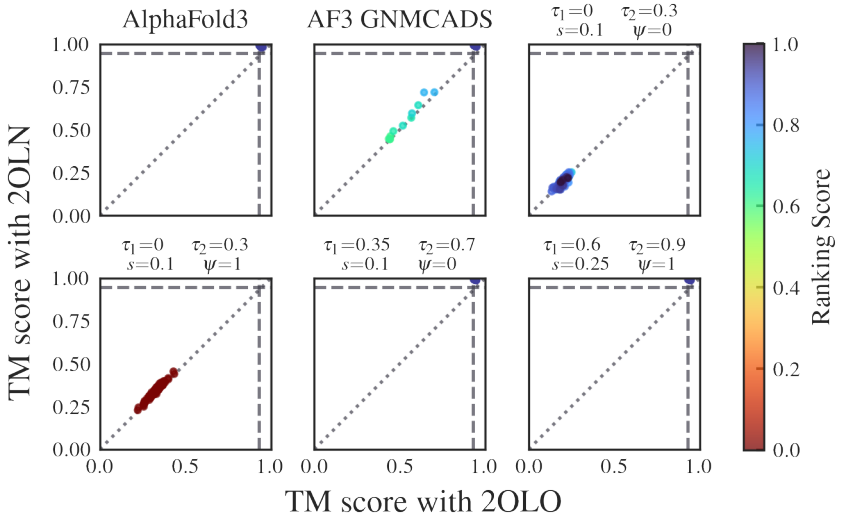
Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



(u) Diversity plots of Q9SS90 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

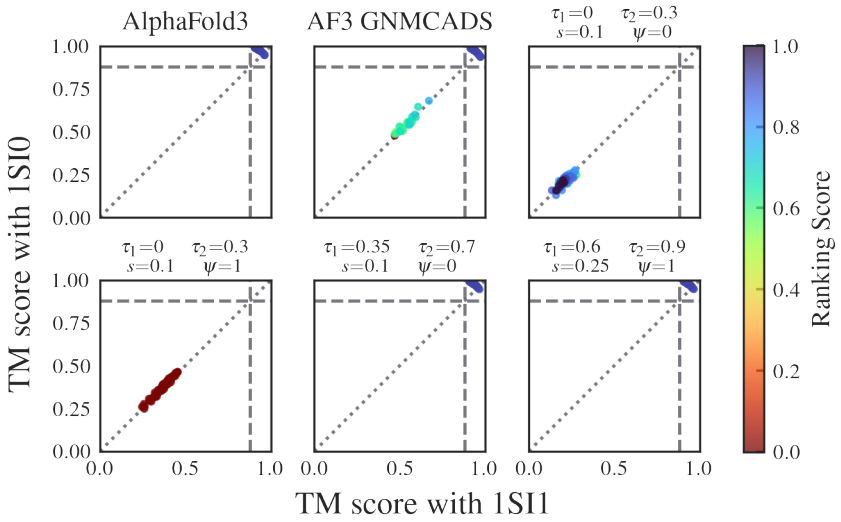


(v) Diversity plots of Q9X6R4 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.



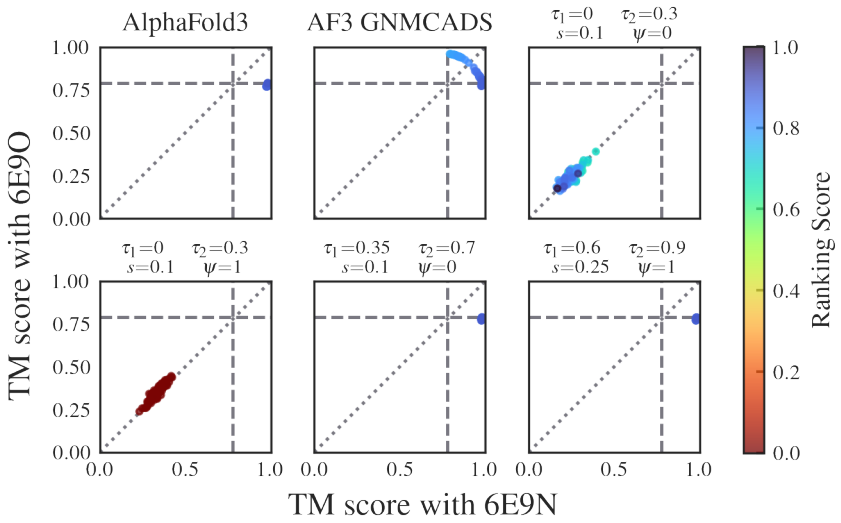
(w) Diversity plots of Q9X9P9 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

Figure 6: Using CADS Without GNM on the Open-Closed dataset (continued on next page).



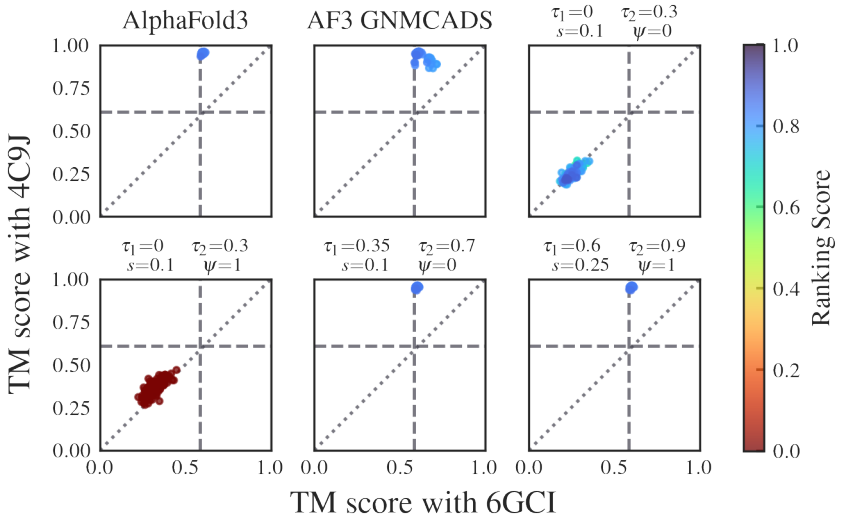
(x) Diversity plots of Q9Z4N6 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

Figure 6: Using CADS Without GNM on the Open-Closed dataset.

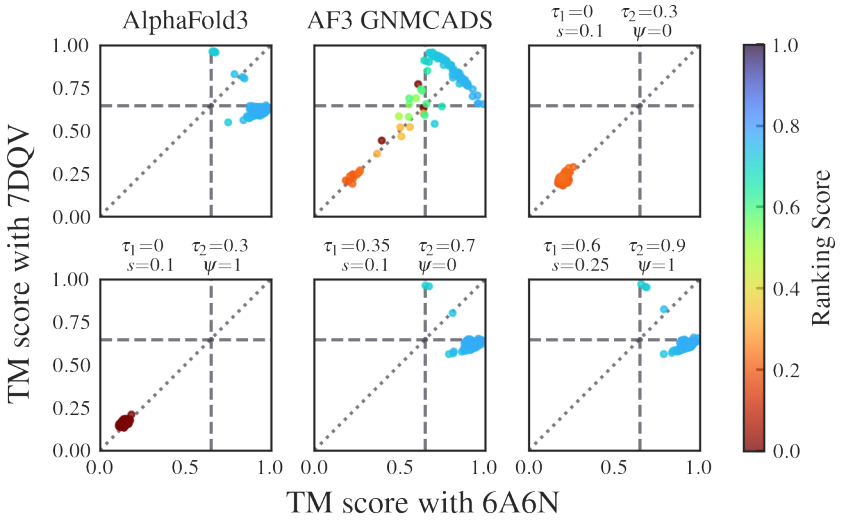


(a) Diversity plots of A5U30_003247 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

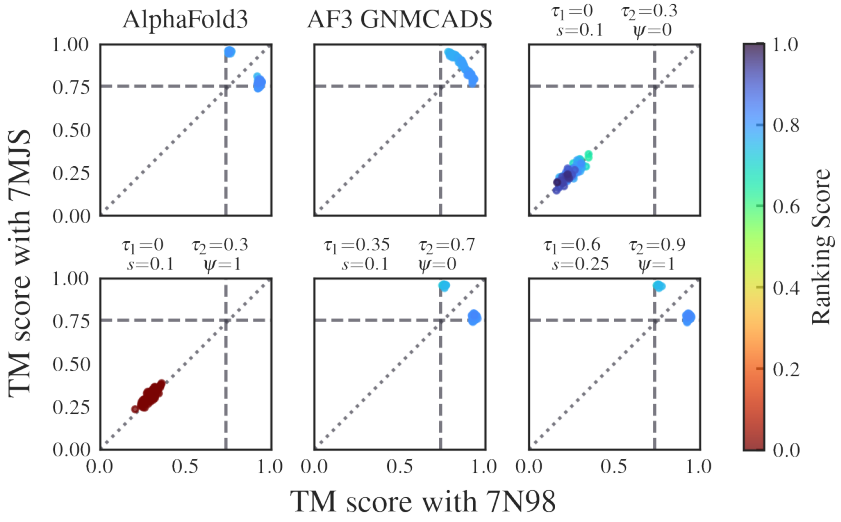
Figure 7: Using CADS Without GNM on the Transporters dataset (continued on next page).



(b) Diversity plots of AAC3 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

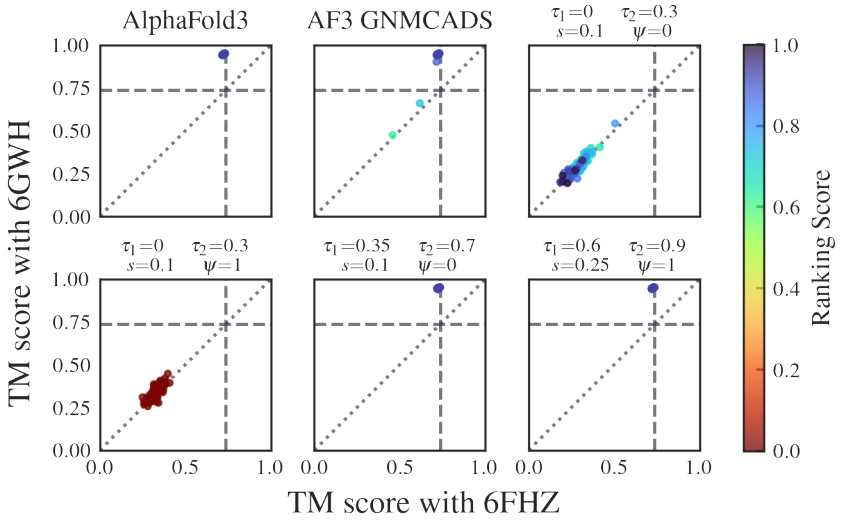


(c) Diversity plots of CYME_CMD148C sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

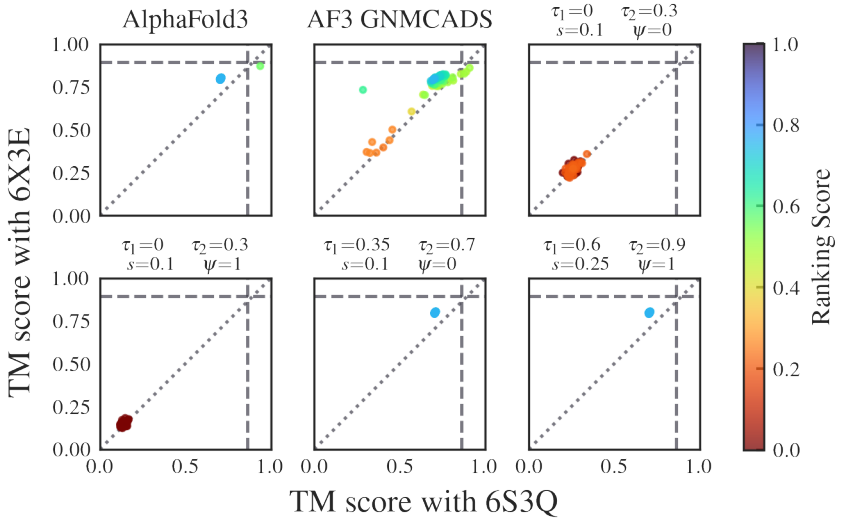


(d) Diversity plots of MFSD2A sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

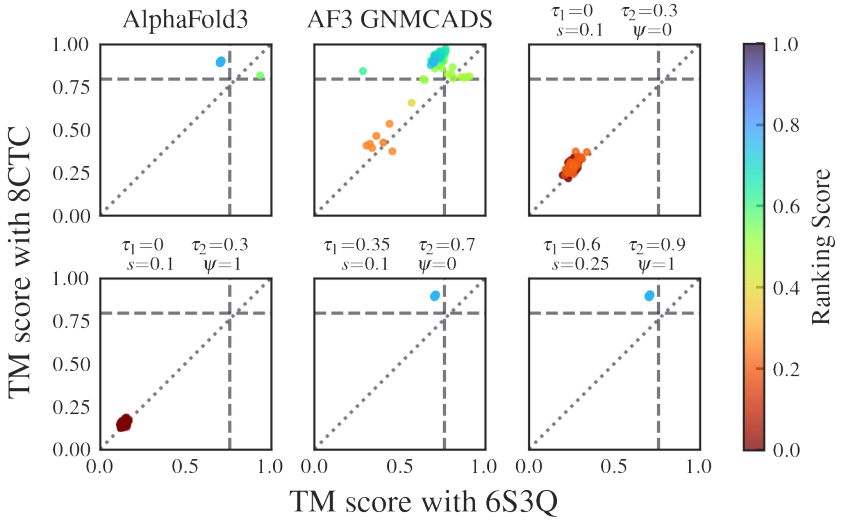
Figure 7: Using CADS Without GNM on the Transporters dataset (continued on next page).



(e) Diversity plots of PF0708 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

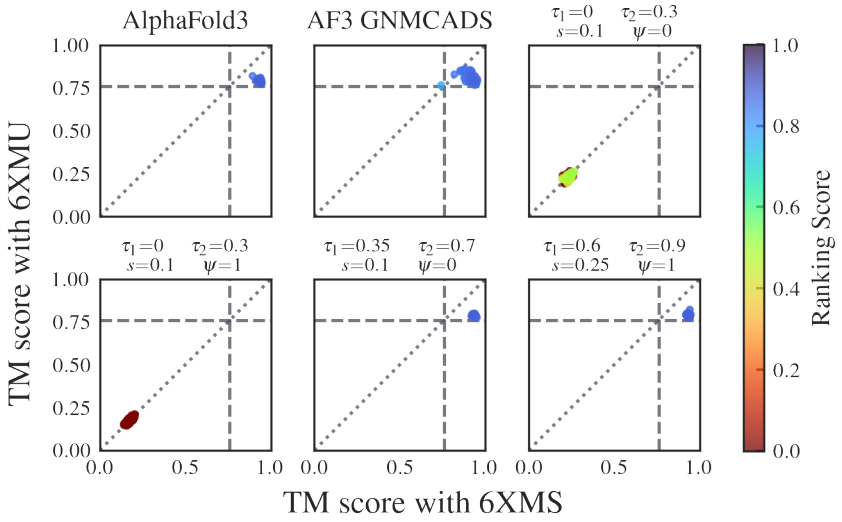


(f) Diversity plots of SLC1A1 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

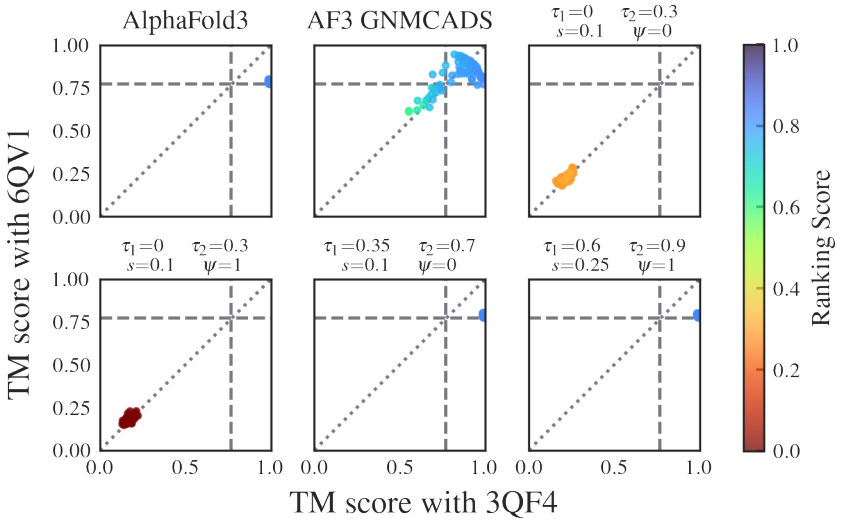


(g) Diversity plots of SLC1A1 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

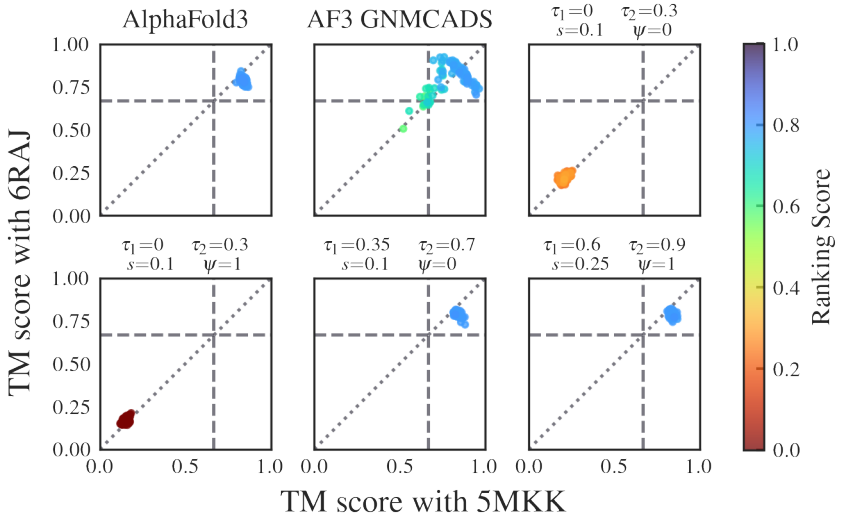
Figure 7: Using CADS Without GNM on the Transporters dataset (continued on next page).



(h) Diversity plots of SPF1 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

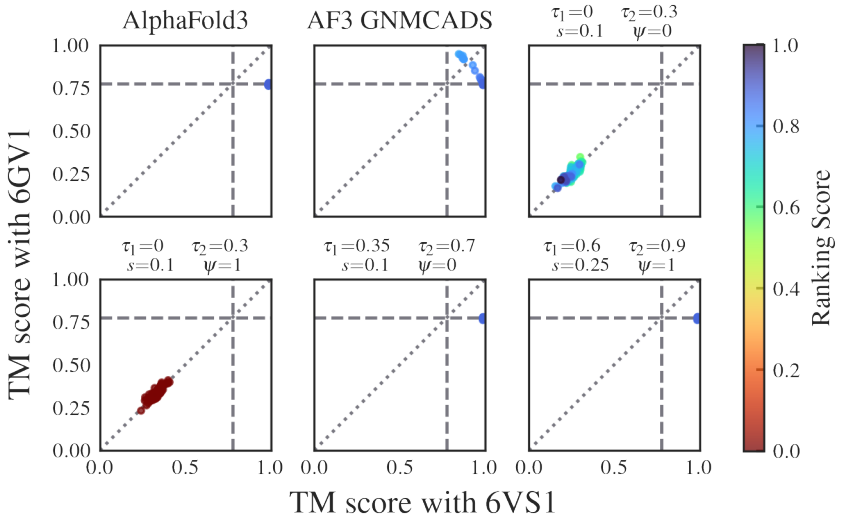


(i) Diversity plots of TM_0287 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

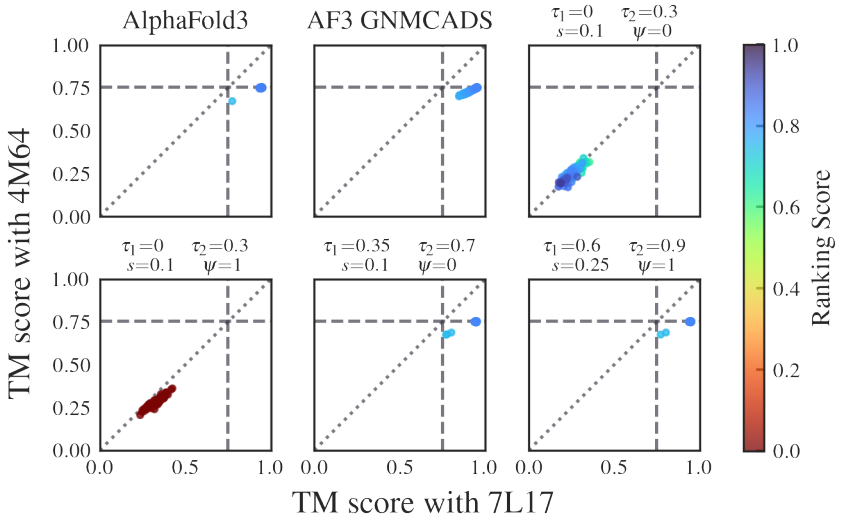


(j) Diversity plots of TT_C0976 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

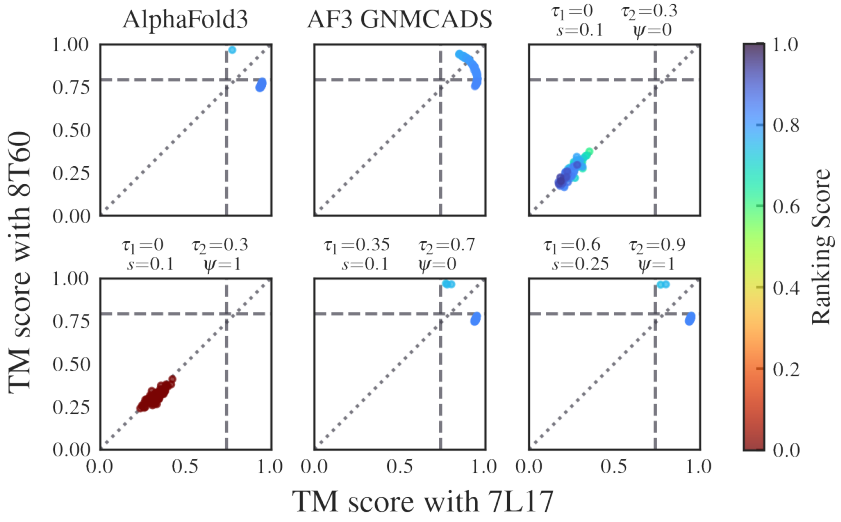
Figure 7: Using CADS Without GNM on the Transporters dataset (continued on next page).



(k) Diversity plots of mdfA sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

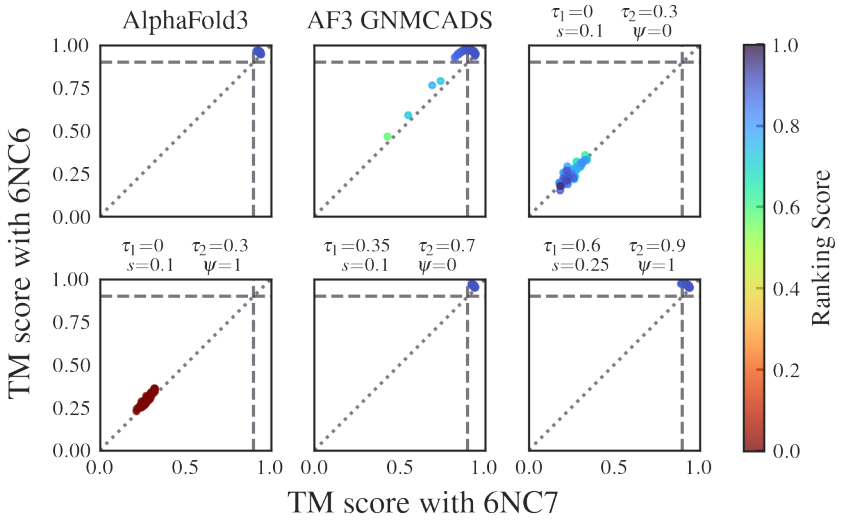


(l) Diversity plots of melB sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

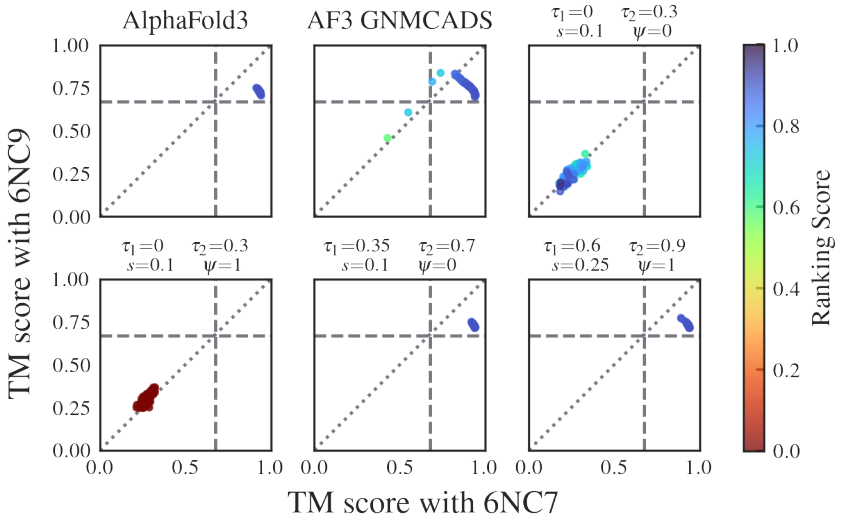


(m) Diversity plots of melB sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

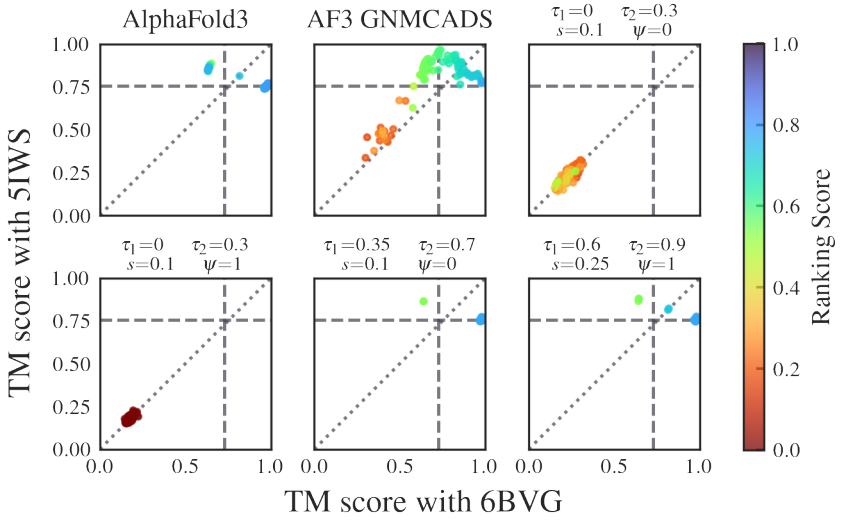
Figure 7: Using CADS Without GNM on the Transporters dataset (continued on next page).



(n) Diversity plots of murJ sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

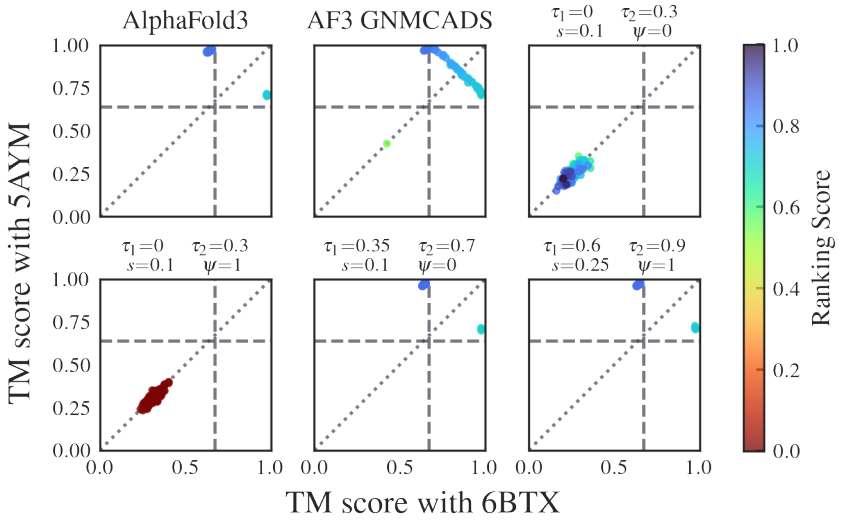


(o) Diversity plots of murJ sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

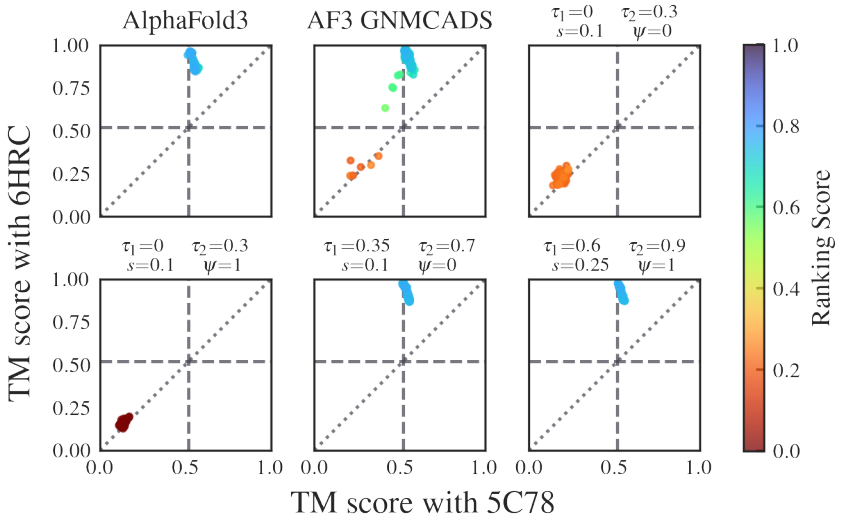


(p) Diversity plots of ptsG sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

Figure 7: Using CADS Without GNM on the Transporters dataset (continued on next page).



(q) Diversity plots of slc39 sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

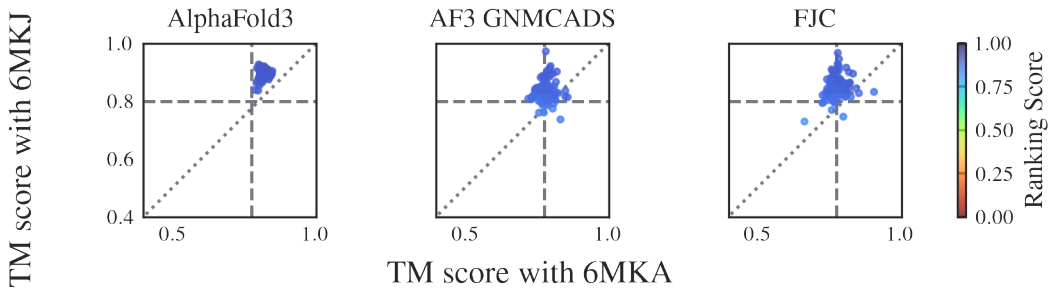


(r) Diversity plots of wlaB sampled with AlphaFold3, AF3 GNMCADS, and AF3 CADS.

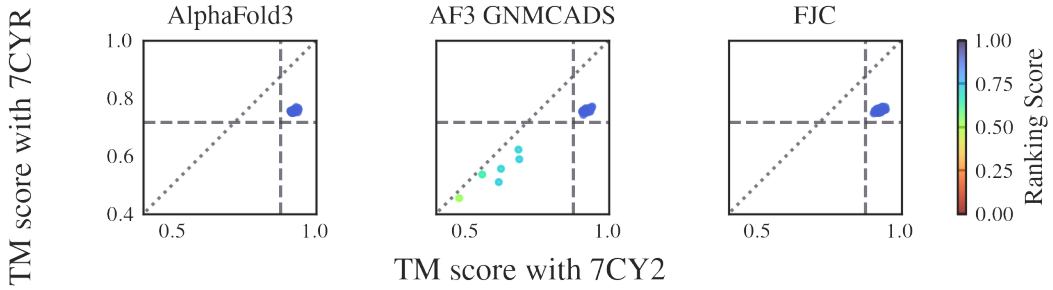
Figure 7: Using CADS Without GNM on the Transporters dataset.

D.2 USING GNMCADS WITH FREELY JOINTED CHAIN

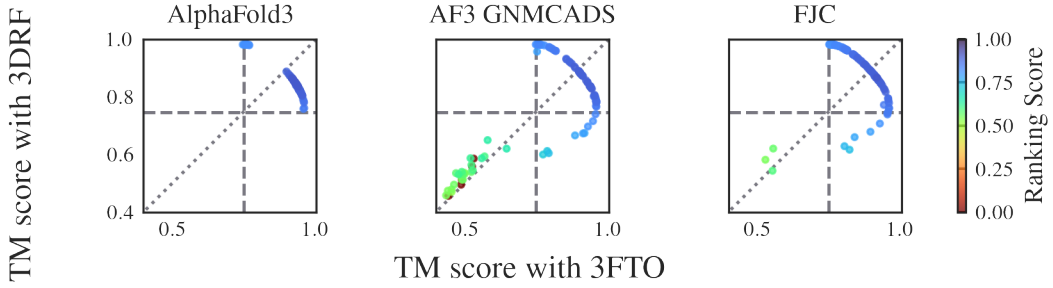
To assess the effect of the protein dynamics, AF3 GNMCADS predictions with the GNM analysis performed on a freely jointed chain (FJC) structure instead of the protein of interest have been performed. Although using dynamical domains based on randomness of FJC structures can still produce wanted results similar to other methods that incorporate randomness into AlphaFold to generate diverse multiple conformations (Wayment-Steele et al., 2024; Stein & Mchaourab, 2022; Kalakoti & Wallner, 2025), as it can be seen in Figures 8l, 8o, 9g, and 9q where incorporating the dynamical properties of the protein of interest with GNM generates more diverse structures.



(a) Diversity plots of A0A075Q0W3 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

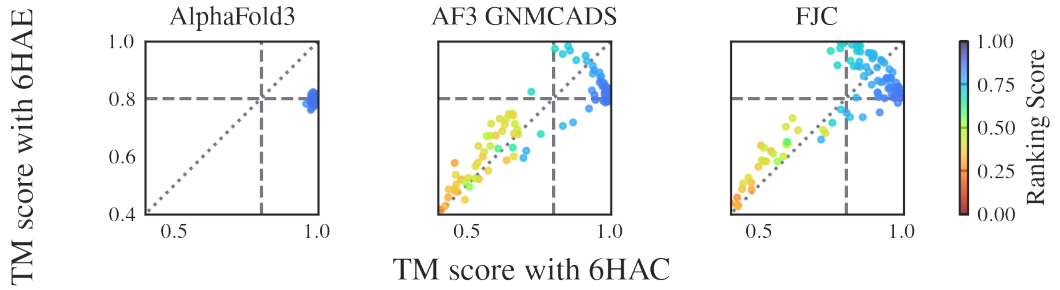


(b) Diversity plots of A0QTT2 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

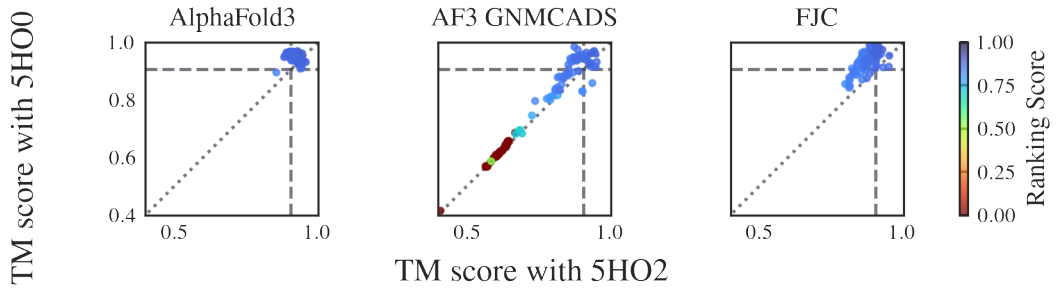


(c) Diversity plots of A2RJ53 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

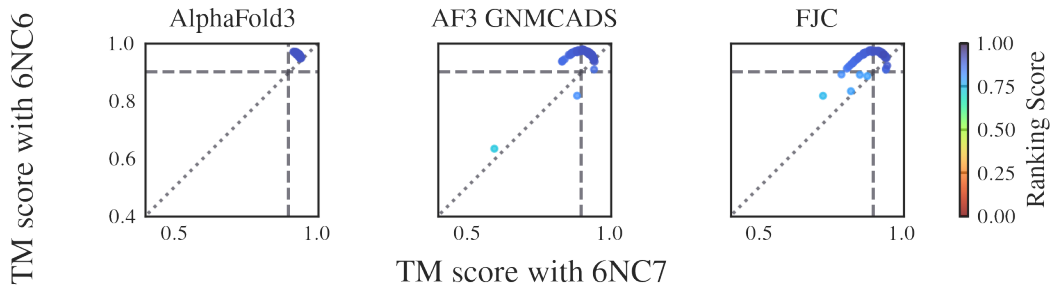
Figure 8: GNMCADS with GNM analysis of FJCs on the Open-Closed dataset (continued on next page).



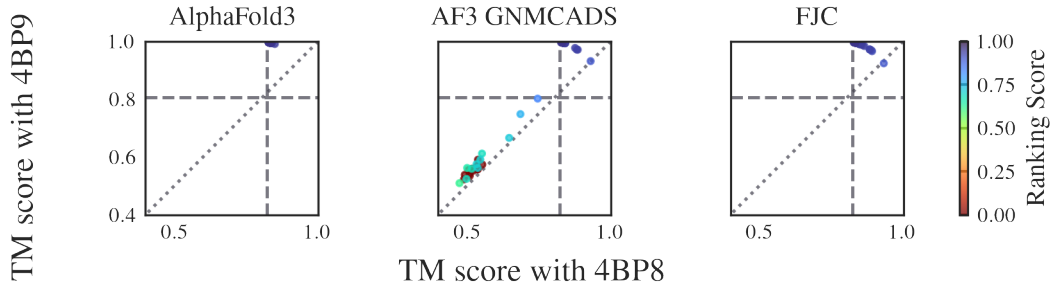
(d) Diversity plots of A6UVT1 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



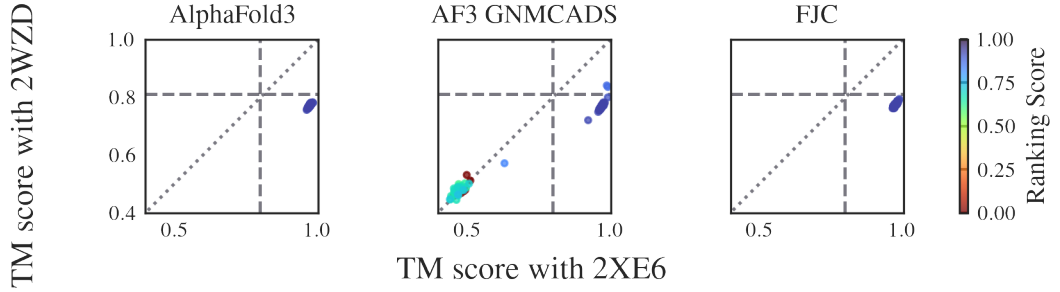
(e) Diversity plots of B3EYN2 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



(f) Diversity plots of B7IE18 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

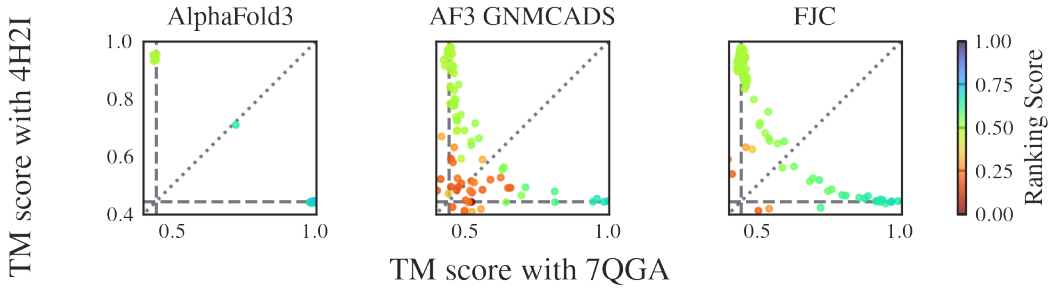


(g) Diversity plots of O76728 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

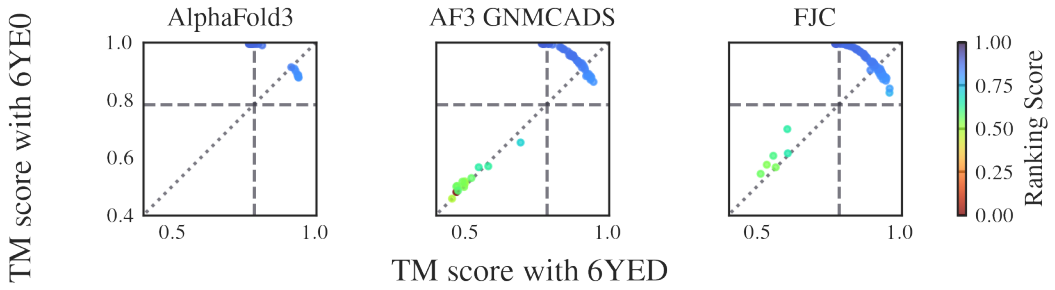


(h) Diversity plots of P00558 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

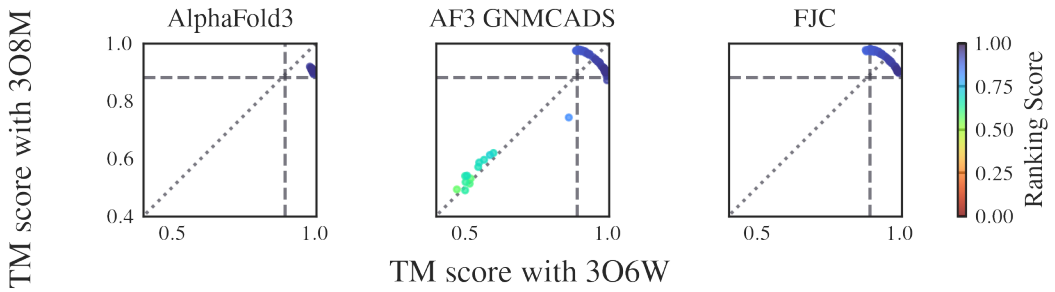
Figure 8: GNMCADS with GNM analysis of FJCs on the Open-Closed dataset (continued on next page).



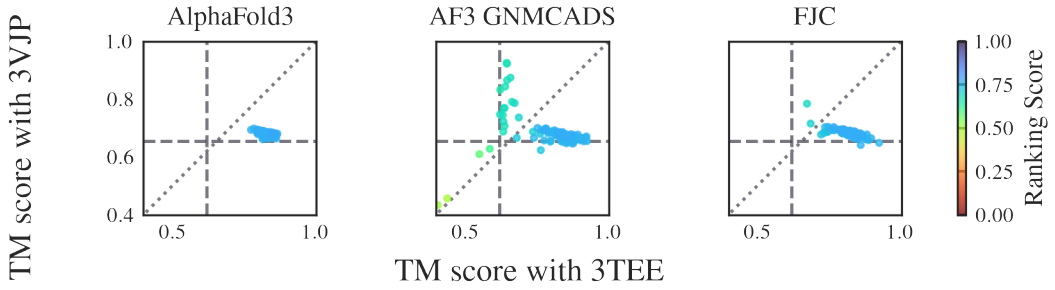
(i) Diversity plots of P21589 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



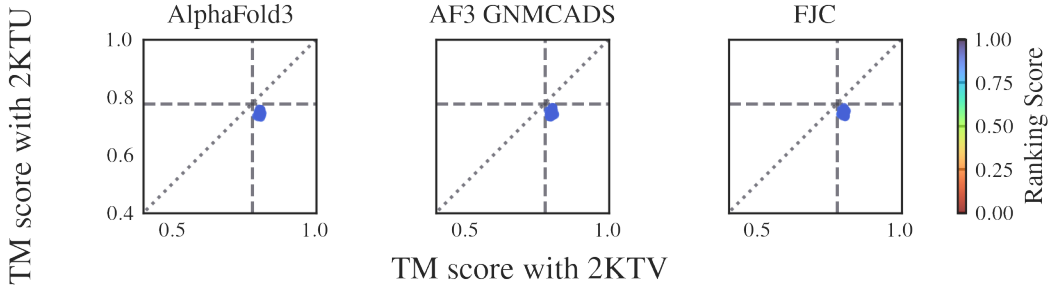
(j) Diversity plots of P31133 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



(k) Diversity plots of P33284 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

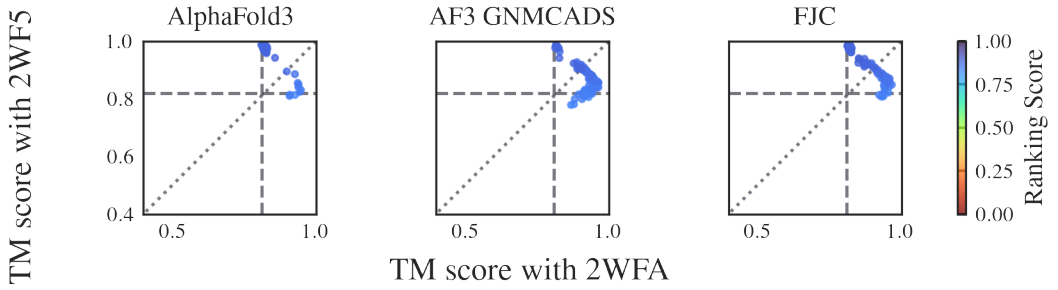


(l) Diversity plots of P40131 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

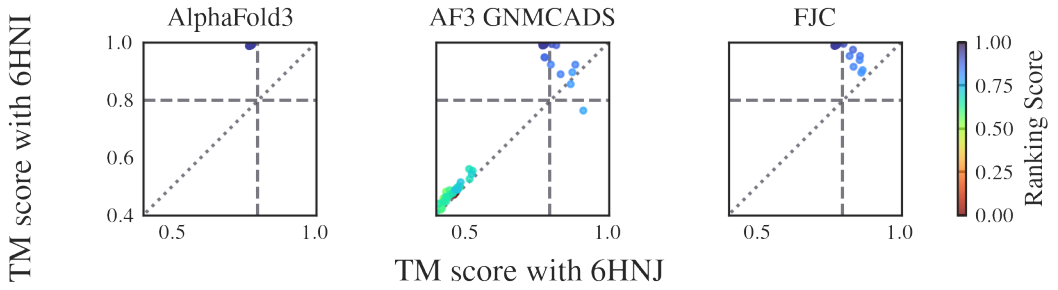


(m) Diversity plots of P62495 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

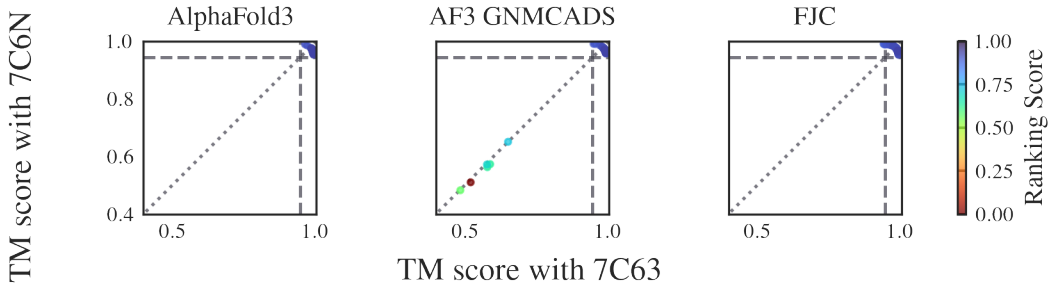
Figure 8: GNMCADS with GNM analysis of FJCs on the Open-Closed dataset (continued on next page).



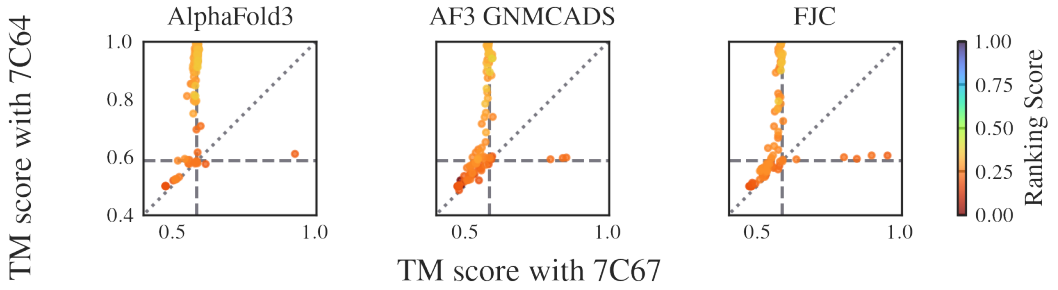
(n) Diversity plots of P71447 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



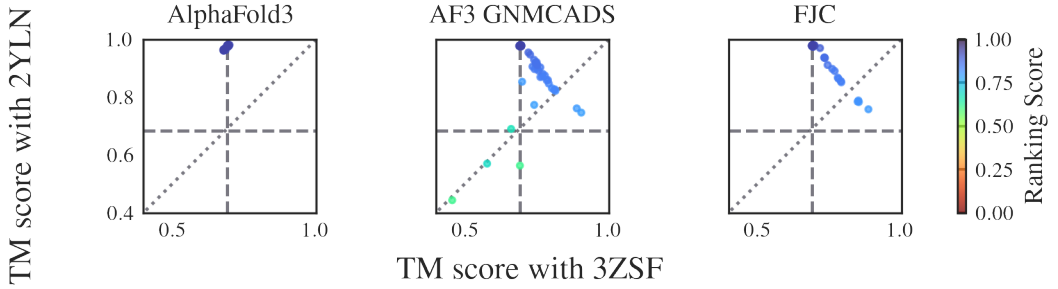
(o) Diversity plots of Q18A65 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



(p) Diversity plots of Q53W80 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

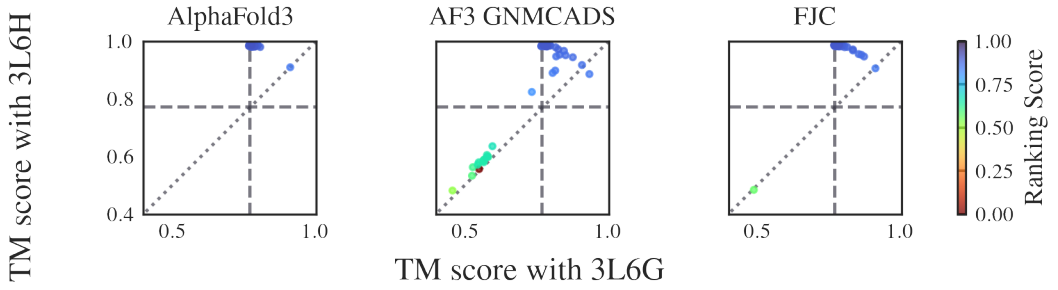


(q) Diversity plots of Q53W80 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

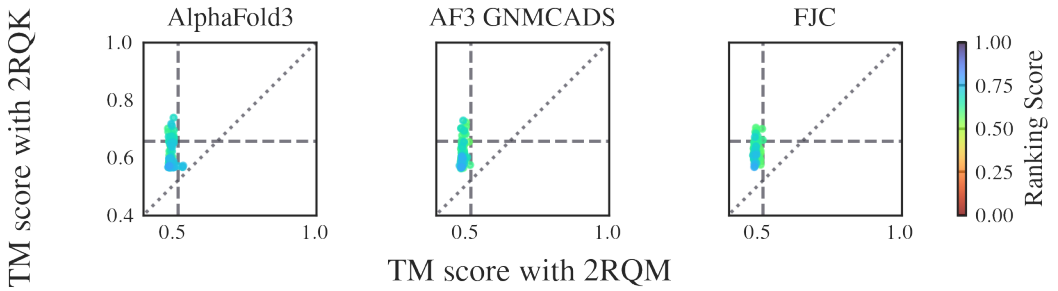


(r) Diversity plots of Q5F9M1 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

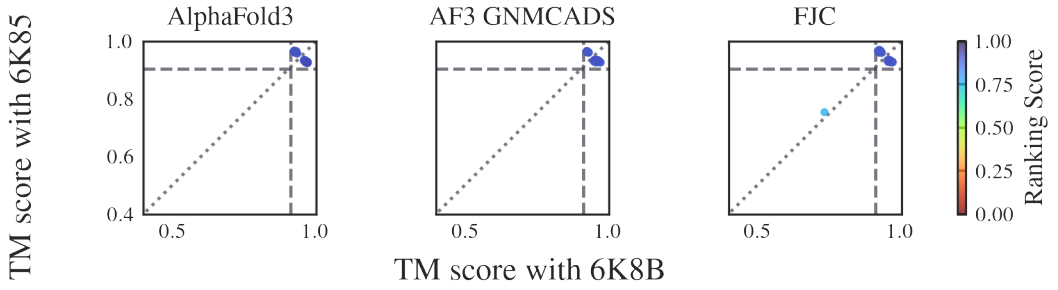
Figure 8: GNMCADS with GNM analysis of FJCs on the Open-Closed dataset (continued on next page).



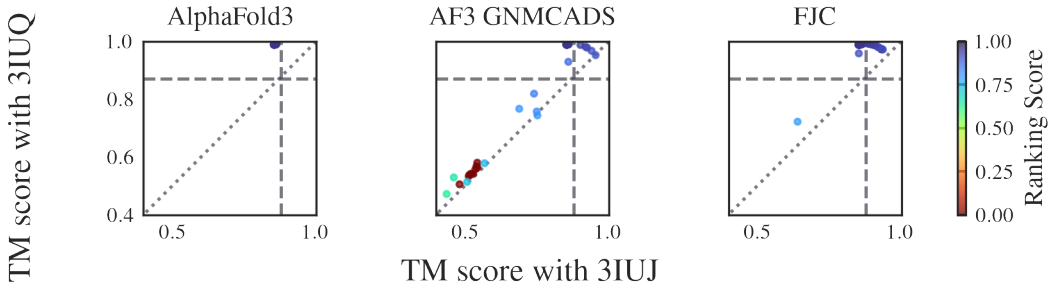
(s) Diversity plots of Q7DAU8 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



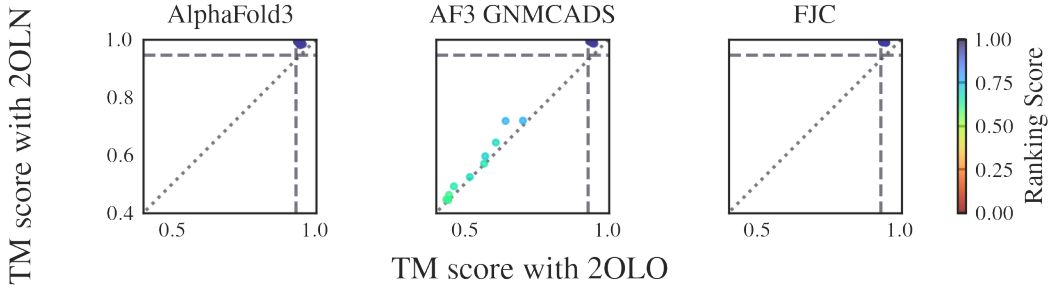
(t) Diversity plots of Q9ERE7 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



(u) Diversity plots of Q9SS90 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

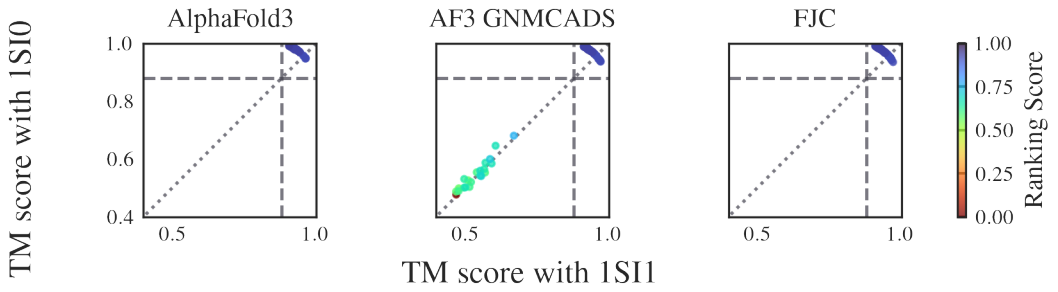


(v) Diversity plots of Q9X6R4 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.



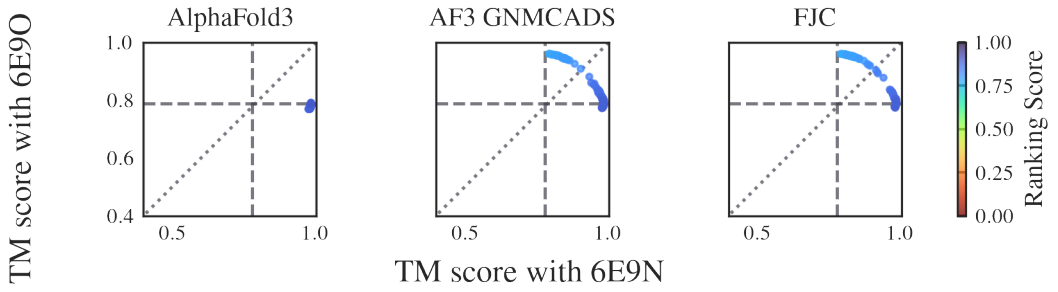
(w) Diversity plots of Q9X9P9 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

Figure 8: GNMCADS with GNM analysis of FJCs on the Open-Closed dataset (continued on next page).

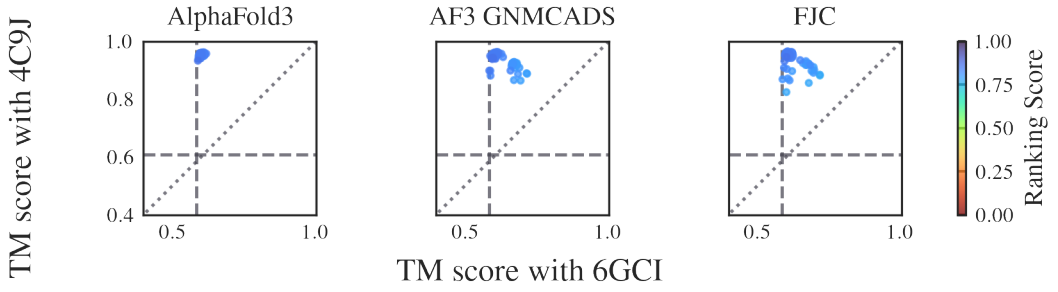


(x) Diversity plots of Q9Z4N6 sampled by AF3, AF3 GNMCADS, and with FJC GNM analysis.

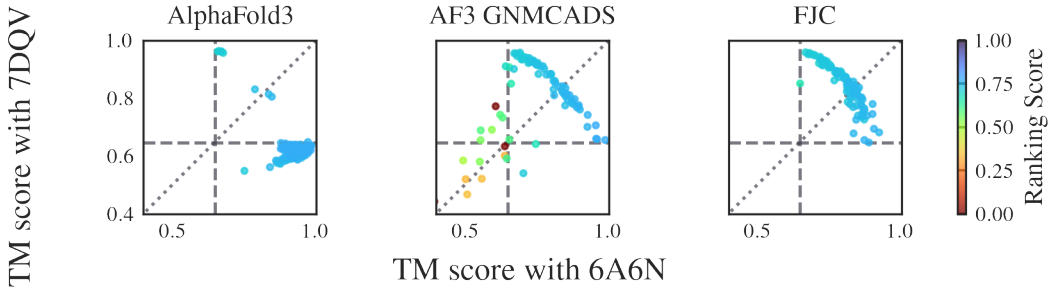
Figure 8: GNMCADS with GNM analysis of FJCs on the Open-Closed dataset.



(a) Diversity plots of A5U30_003247 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

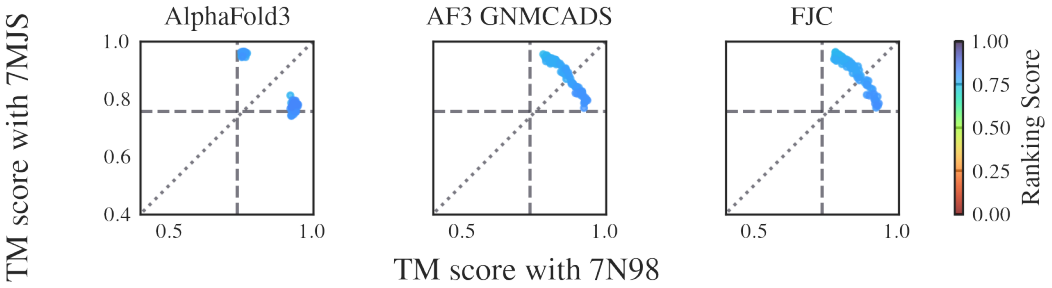


(b) Diversity plots of AAC3 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

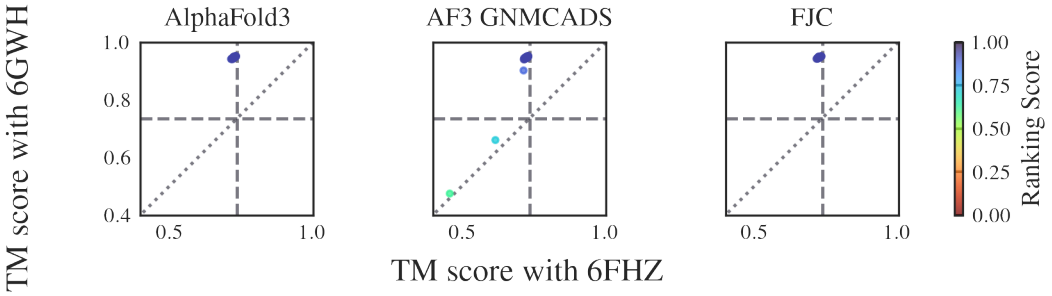


(c) Diversity plots of CYME_CMD148C sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

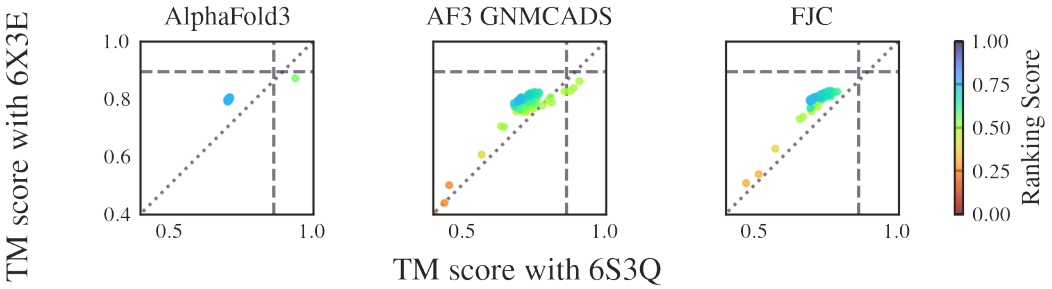
Figure 9: GNMCADS with GNM analysis of FJC on the Transporters dataset (continued on next page).



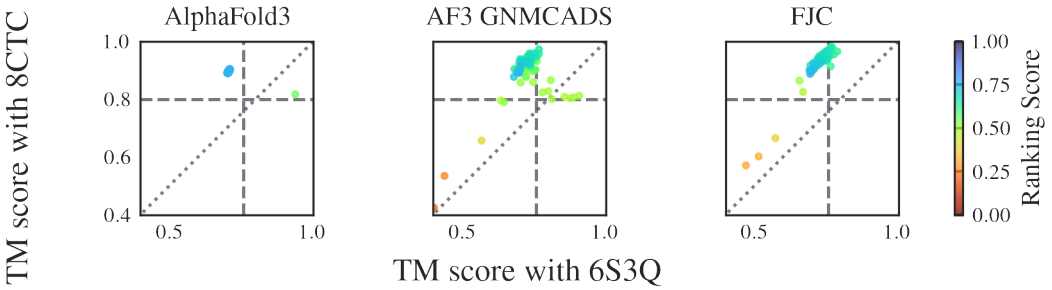
(d) Diversity plots of MFSD2A sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.



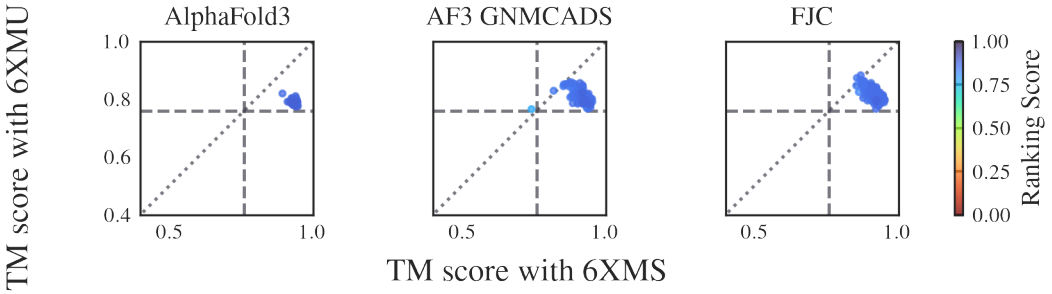
(e) Diversity plots of PF0708 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.



(f) Diversity plots of SLC1A1 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

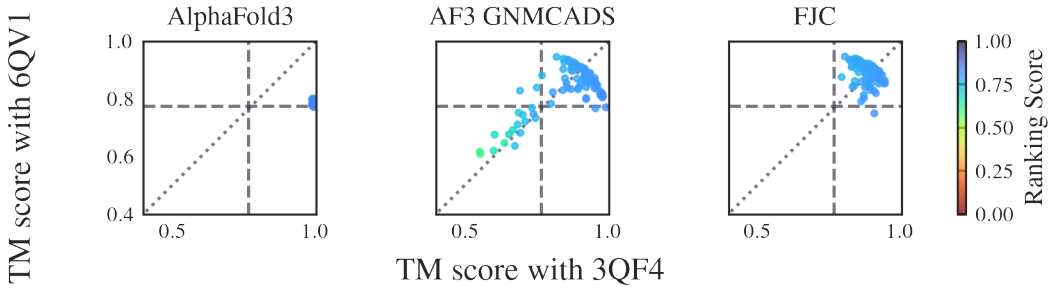


(g) Diversity plots of SLC1A1 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

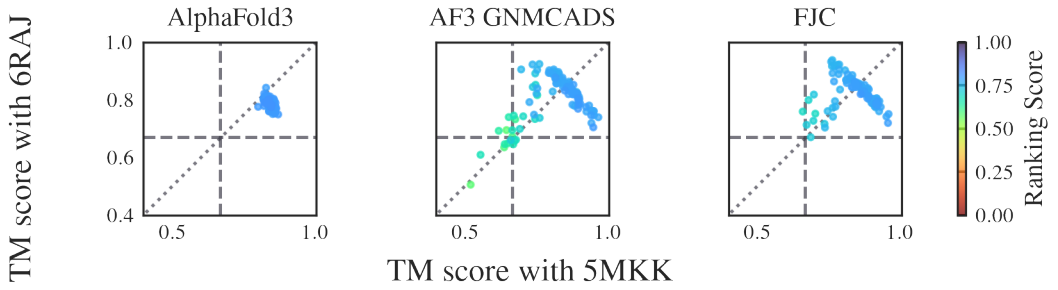


(h) Diversity plots of SPF1 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

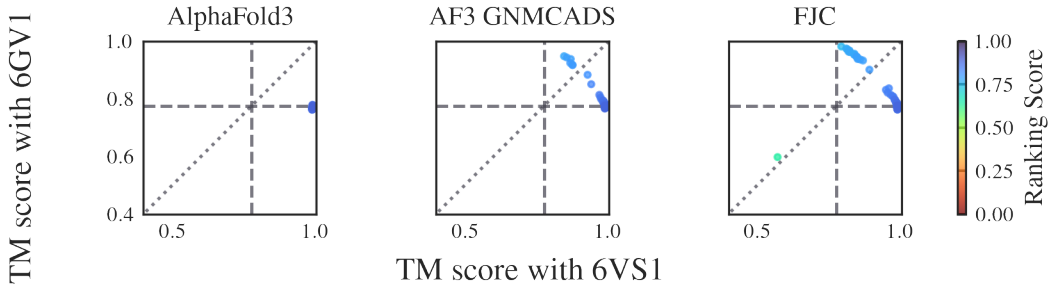
Figure 9: GNMCADS with GNM analysis of FJC on the Transporters dataset (continued on next page).



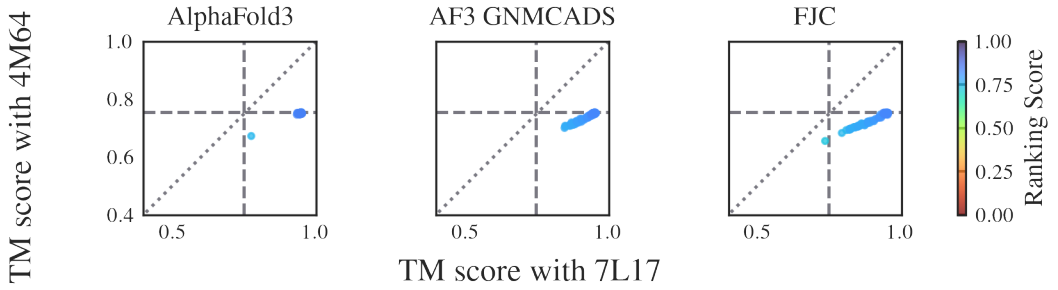
(i) Diversity plots of TM_0287 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.



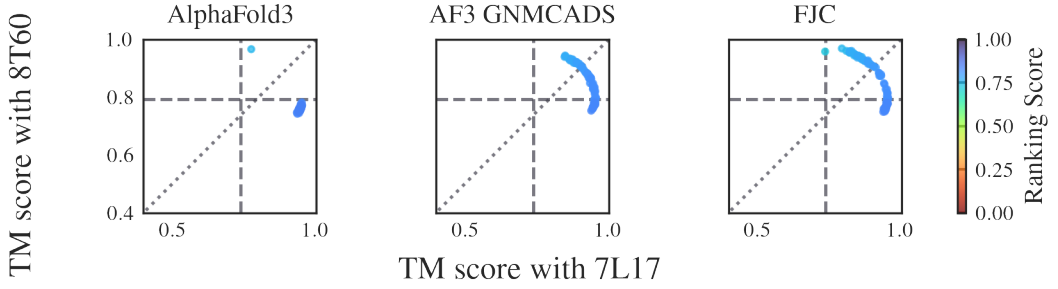
(j) Diversity plots of TT_C0976 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.



(k) Diversity plots of mdfA sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

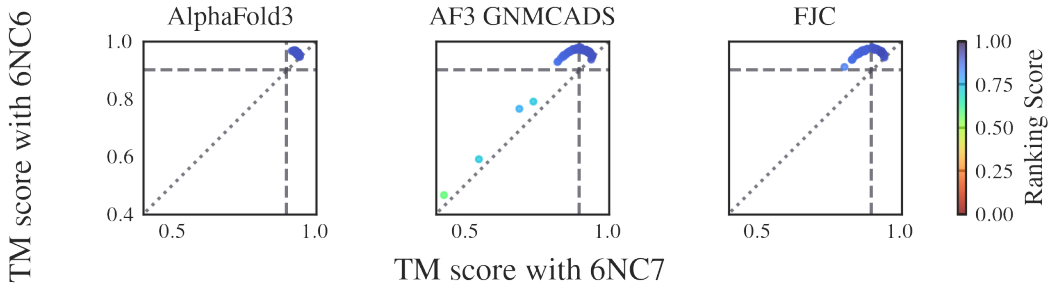


(l) Diversity plots of melB sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

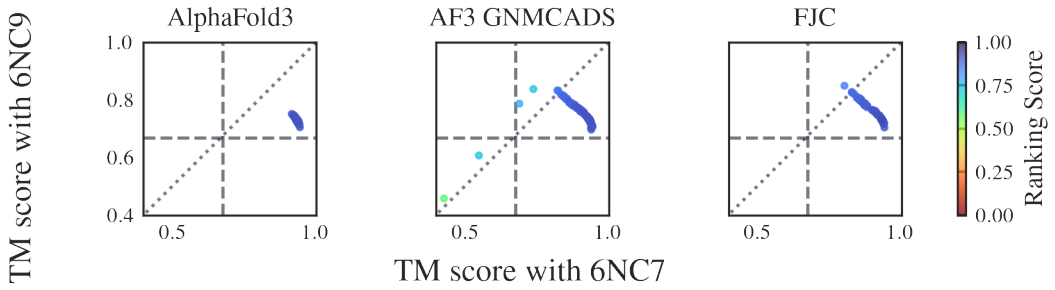


(m) Diversity plots of melB sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

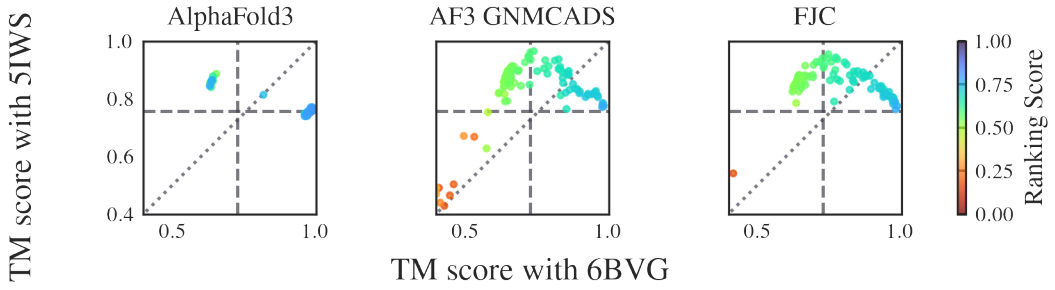
Figure 9: GNMCADS with GNM analysis of FJC on the Transporters dataset (continued on next page).



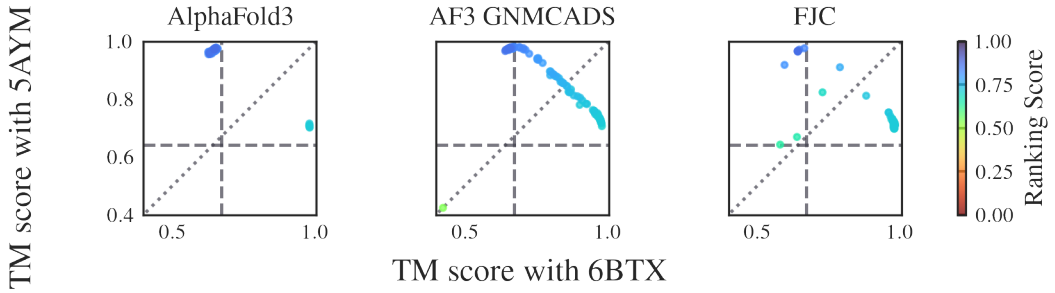
(n) Diversity plots of murJ sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.



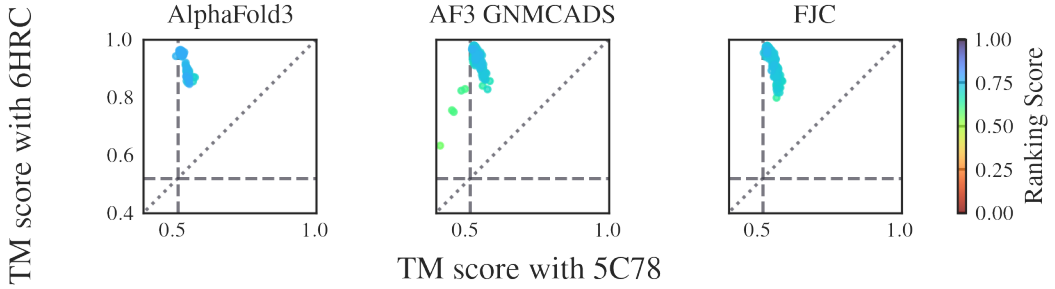
(o) Diversity plots of murJ sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.



(p) Diversity plots of ptsG sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.



(q) Diversity plots of slc39 sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

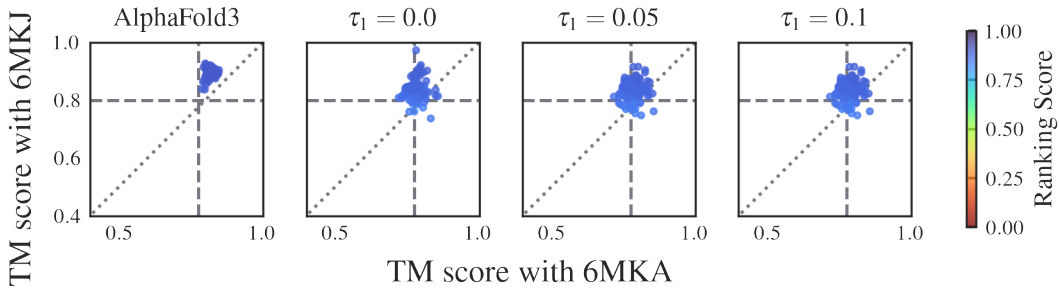


(r) Diversity plots of wlaB sampled with AlphaFold3, AF3 GNMCADS, and with FJC GNM analysis.

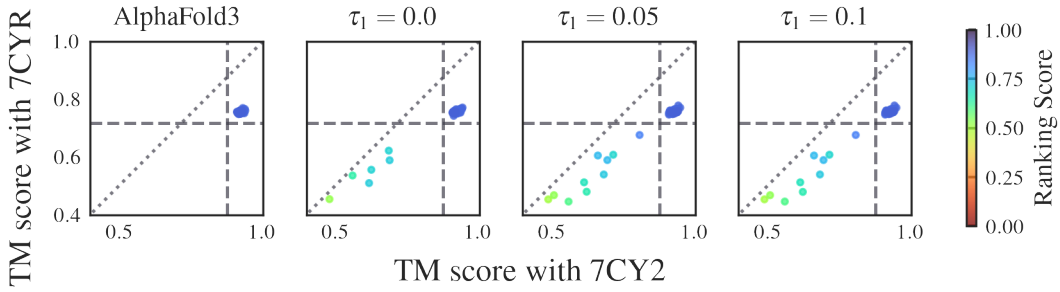
Figure 9: GNMCADS with GNM analysis of FJC on the Transporters dataset.

D.3 VARYING THE VALUE OF τ_1

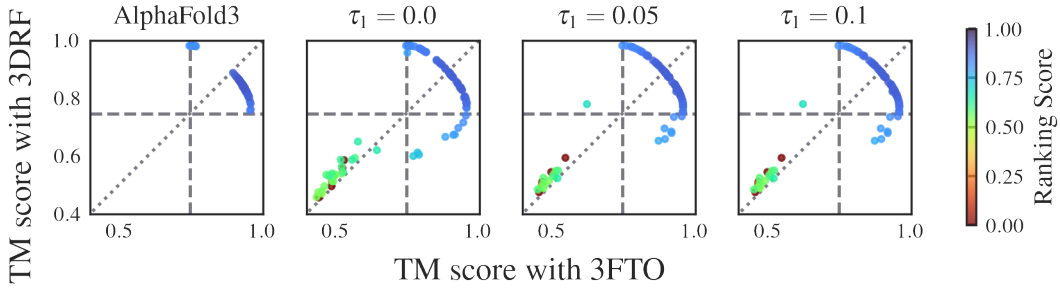
The value of τ_1 represents the stop point for annealing for GNMCADS where the annealing no corruption is applied for the values $t \leq tau_1$. A value of 0 for τ_1, which is used by default in AF3 GNMCADS, ensures that the condition annealing will not be cut off and will applied at each diffusion step.



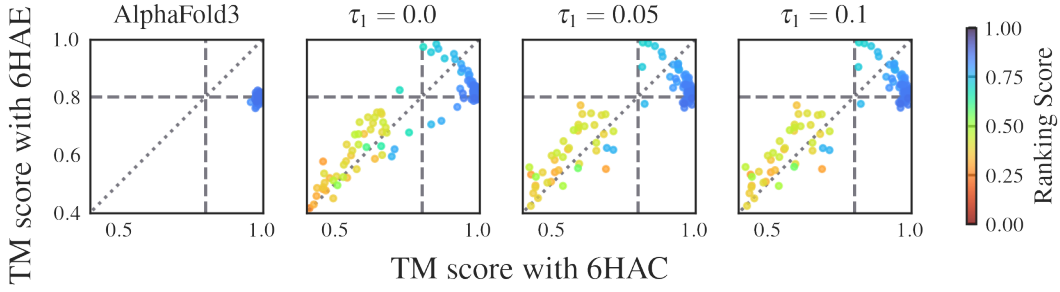
(a) Diversity plots of A0A075Q0W3 for different τ_1 values.



(b) Diversity plots of A0QTT2 for different τ_1 values.

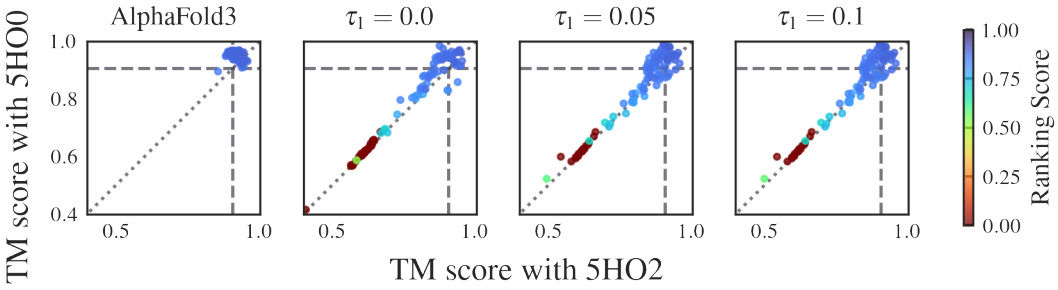


(c) Diversity plots of A2RJ53 for different τ_1 values.

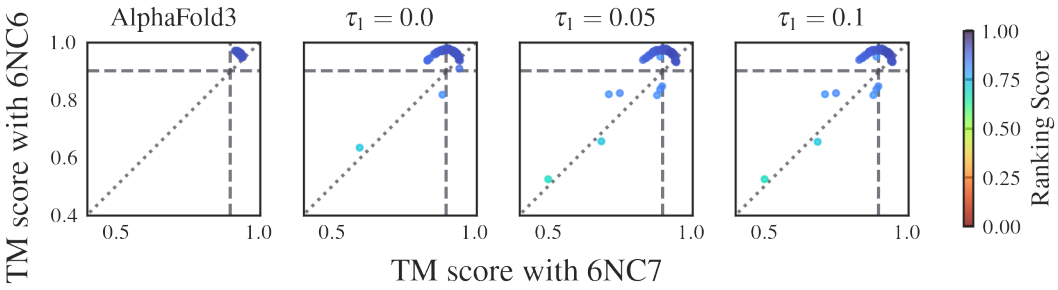


(d) Diversity plots of A6UVT1 for different τ_1 values.

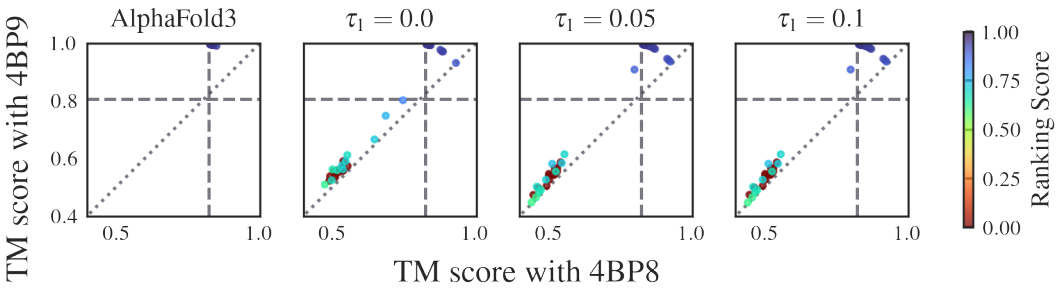
Figure 10: Effect of the CADS parameter τ_1 on the Open-Closed dataset (continued on next page).



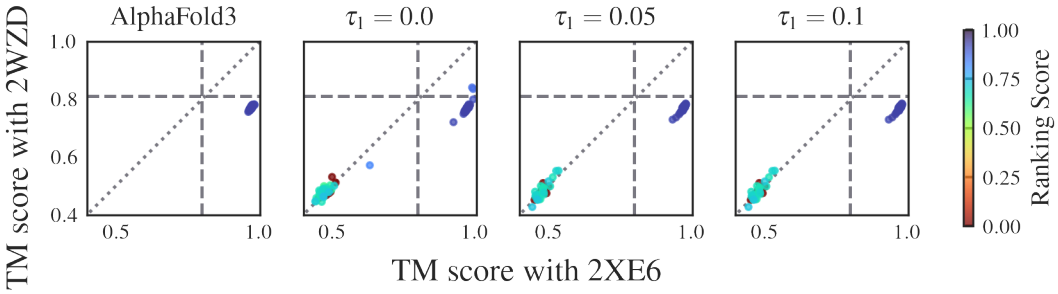
(e) Diversity plots of B3EYN2 for different τ_1 values.



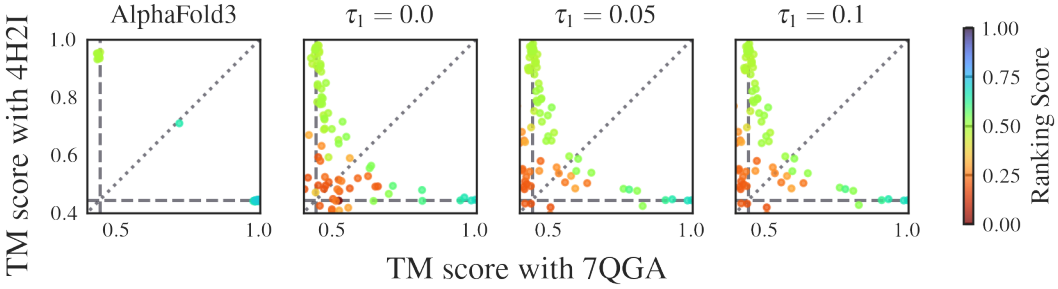
(f) Diversity plots of B7IE18 for different τ_1 values.



(g) Diversity plots of O76728 for different τ_1 values.

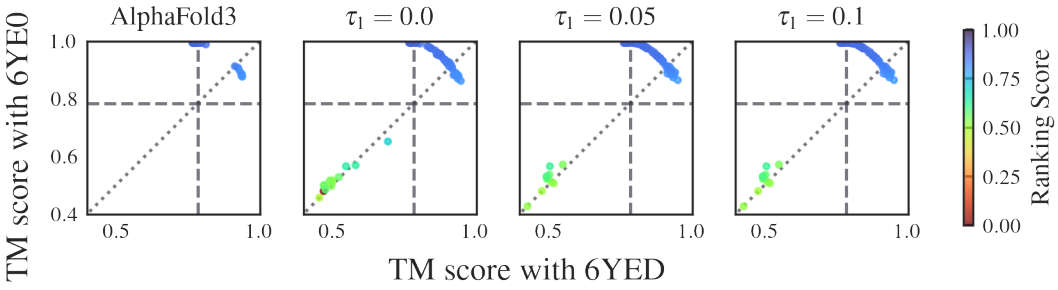


(h) Diversity plots of P00558 for different τ_1 values.

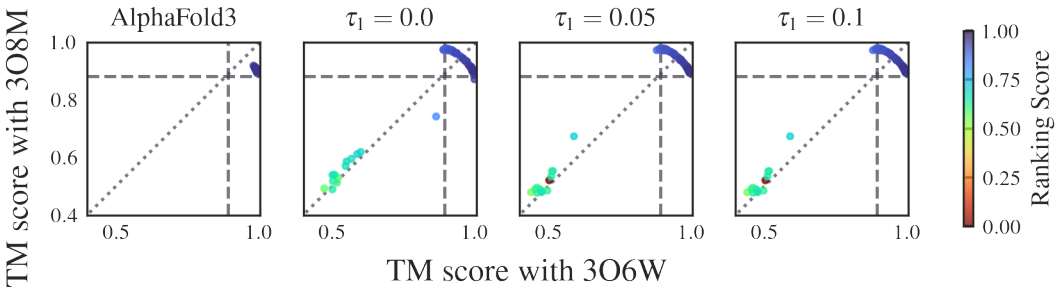


(i) Diversity plots of P21589 for different τ_1 values.

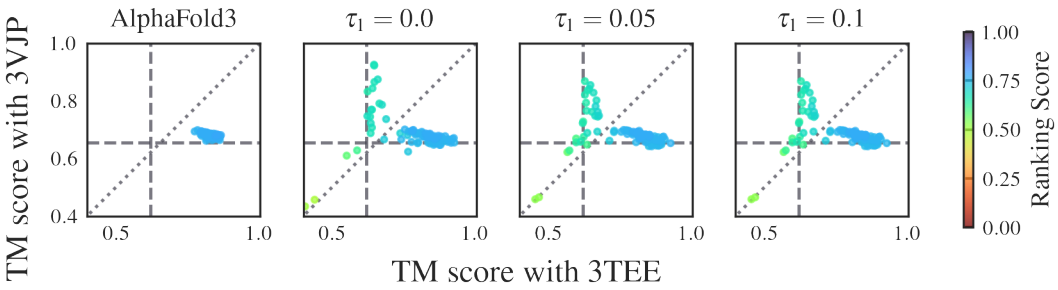
Figure 10: Effect of the CADS parameter τ_1 on the Open-Closed dataset (continued on next page).



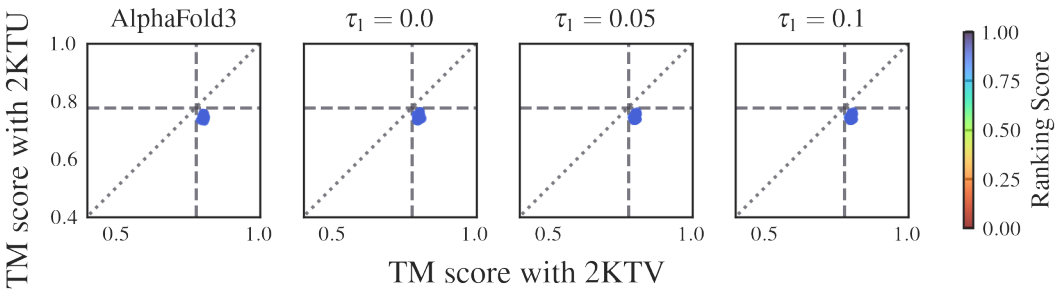
(j) Diversity plots of P31133 for different τ_1 values.



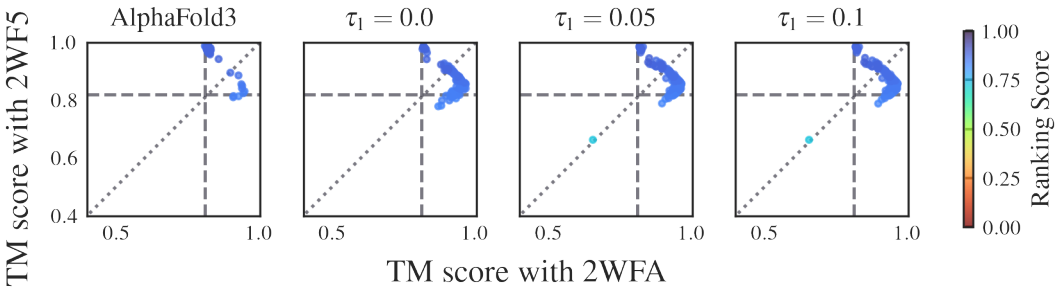
(k) Diversity plots of P33284 for different τ_1 values.



(l) Diversity plots of P40131 for different τ_1 values.

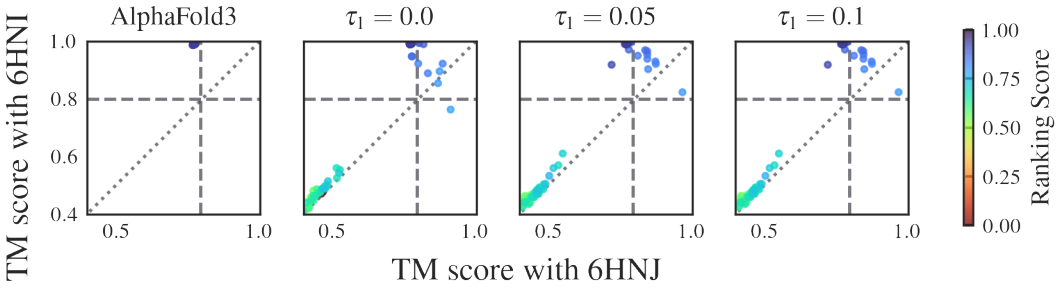


(m) Diversity plots of P62495 for different τ_1 values.

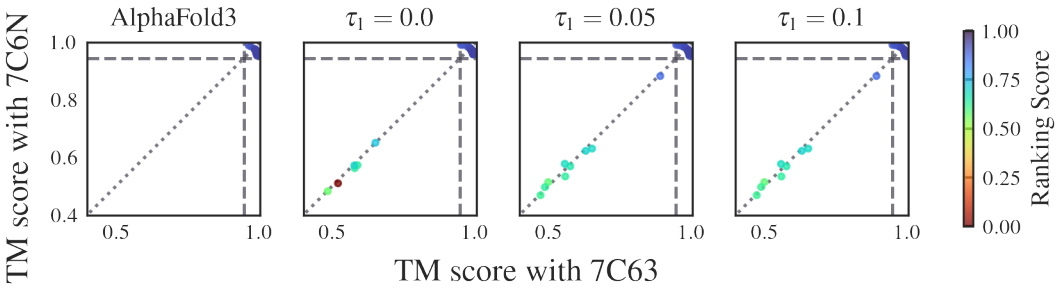


(n) Diversity plots of P71447 for different τ_1 values.

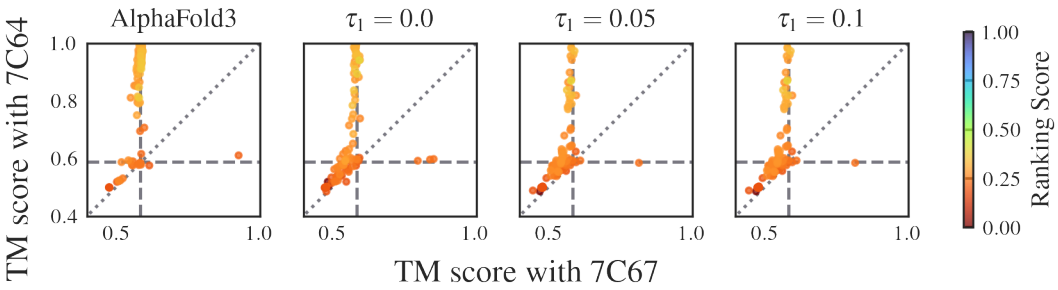
Figure 10: Effect of the CADS parameter τ_1 on the Open-Closed dataset (continued on next page).



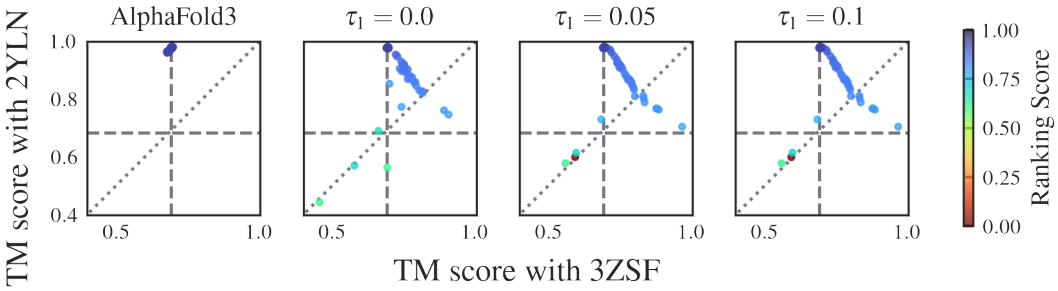
(o) Diversity plots of Q18A65 for different τ_1 values.



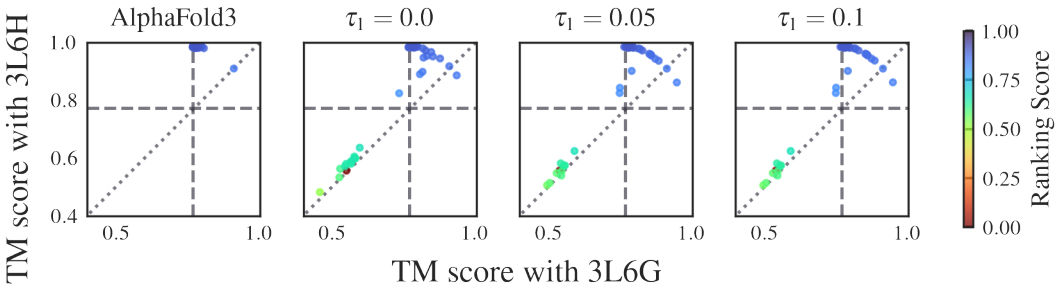
(p) Diversity plots of Q53W80 for different τ_1 values.



(q) Diversity plots of Q53W80 for different τ_1 values.

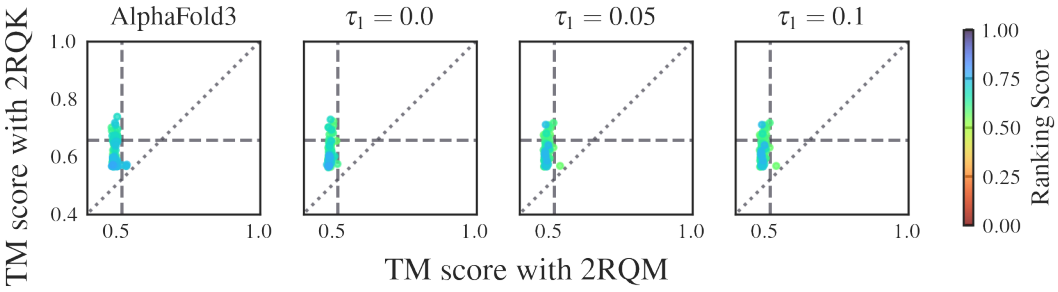


(r) Diversity plots of Q5F9M1 for different τ_1 values.

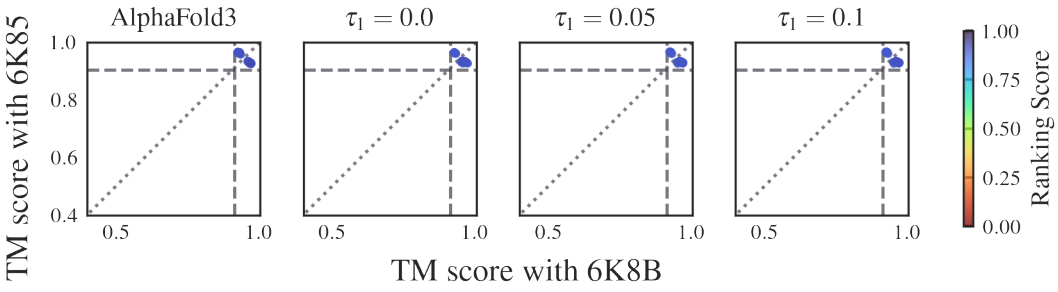


(s) Diversity plots of Q7DAU8 for different τ_1 values.

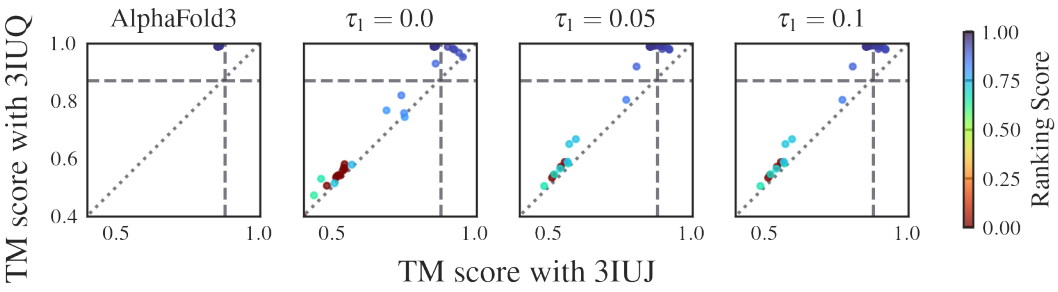
Figure 10: Effect of the CADS parameter τ_1 on the Open-Closed dataset (continued on next page).



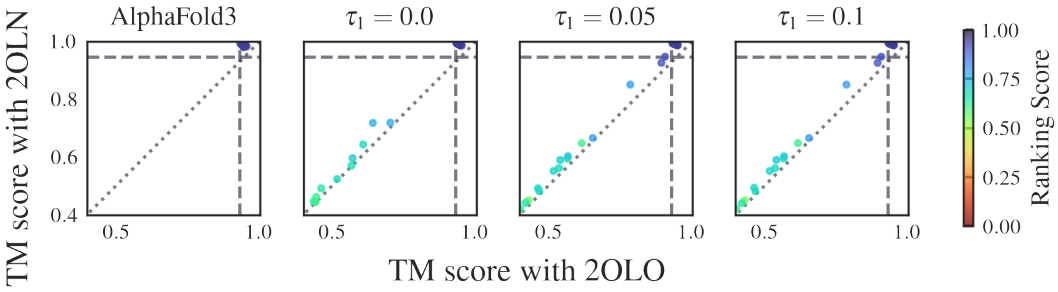
(t) Diversity plots of Q9ERE7 for different τ_1 values.



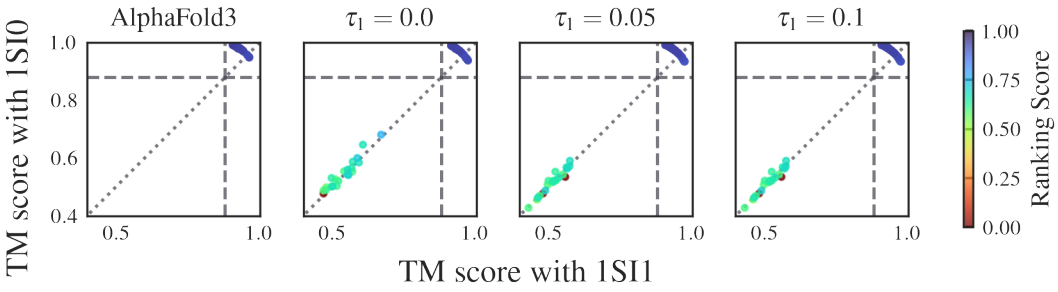
(u) Diversity plots of Q9SS90 for different τ_1 values.



(v) Diversity plots of Q9X6R4 for different τ_1 values.

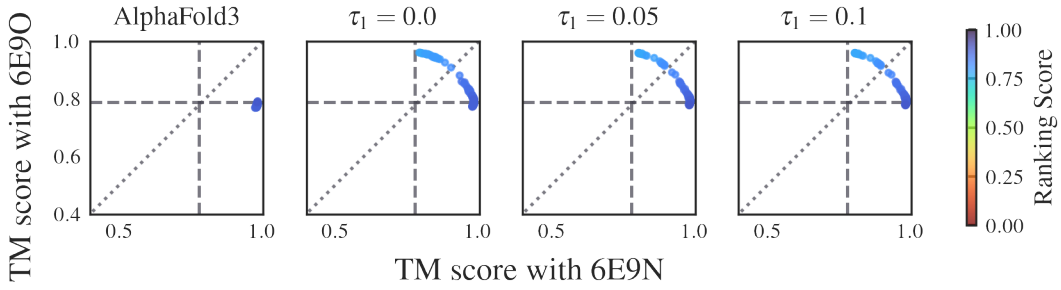


(w) Diversity plots of Q9X9P9 for different τ_1 values.

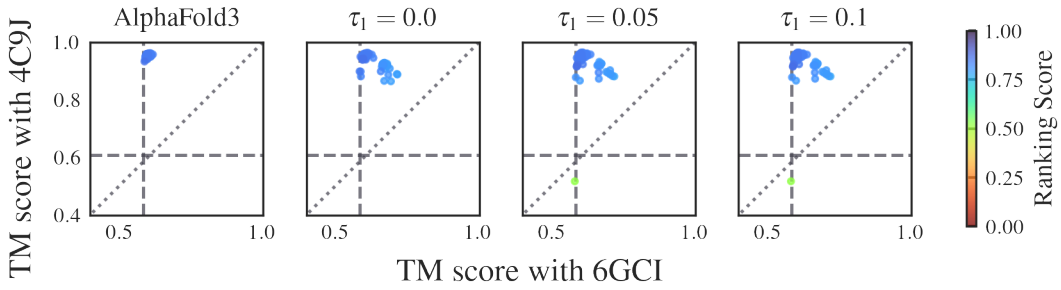


(x) Diversity plots of Q9Z4N6 for different τ_1 values.

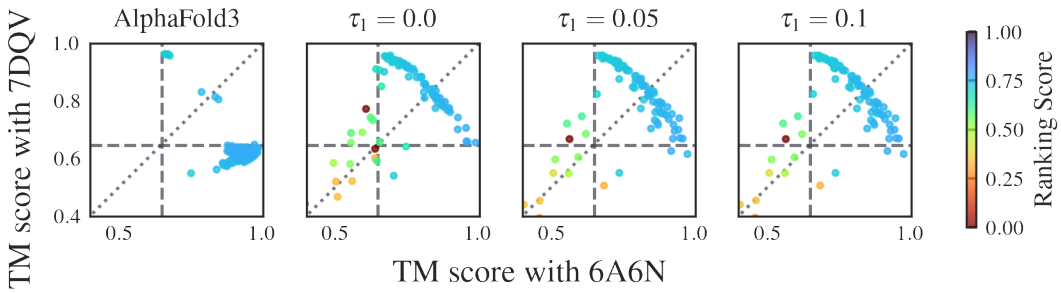
Figure 10: Effect of the CADS parameter τ_1 on the Open-Closed dataset.



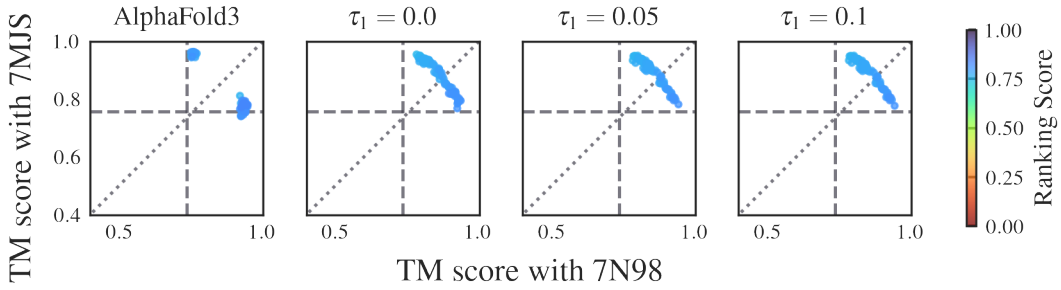
(a) Diversity plots of A5U30_003247 for different τ_1 values.



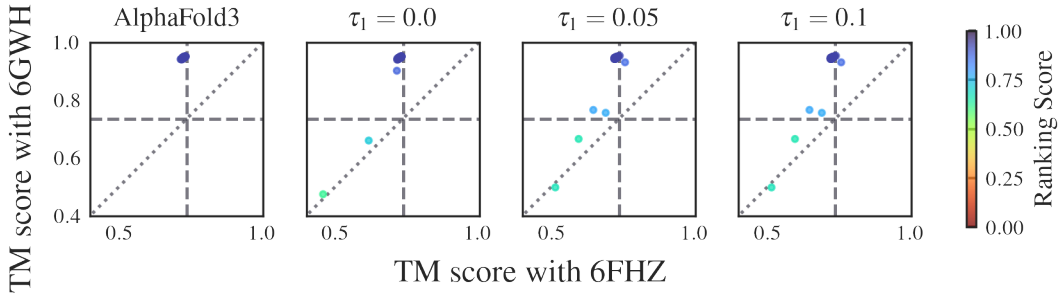
(b) Diversity plots of AAC3 for different τ_1 values.



(c) Diversity plots of CYME_CMD148C for different τ_1 values.

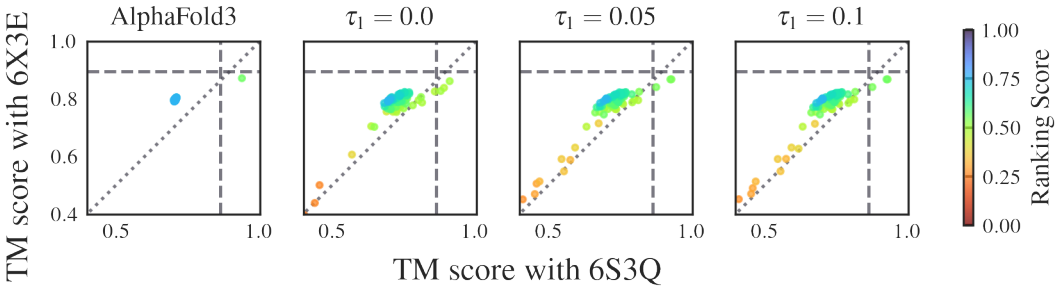


(d) Diversity plots of MFSD2A for different τ_1 values.

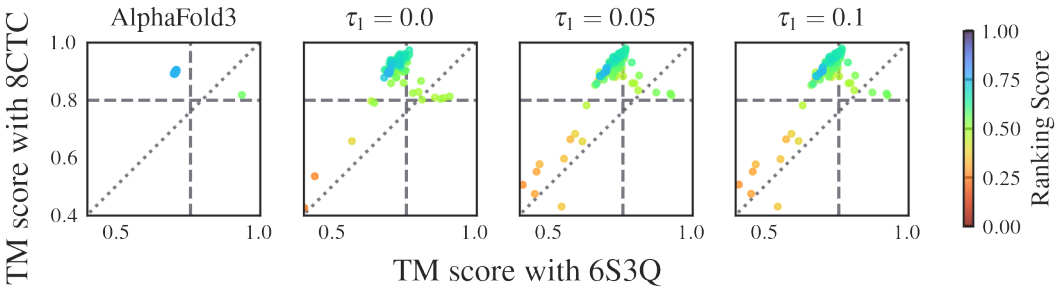


(e) Diversity plots of PF0708 for different τ_1 values.

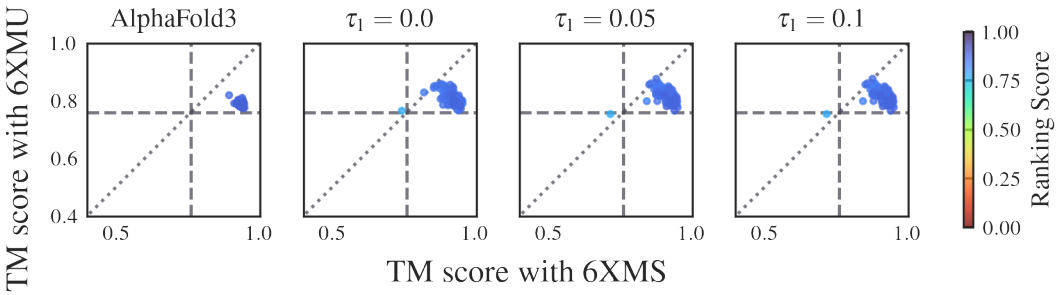
Figure 11: Effect of the CADS parameter τ_1 on the Transporters dataset (continued on next page).



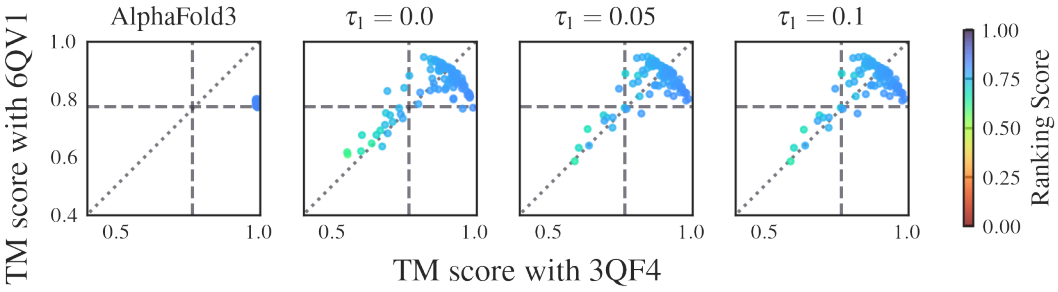
(f) Diversity plots of SLC1A1 for different τ_1 values.



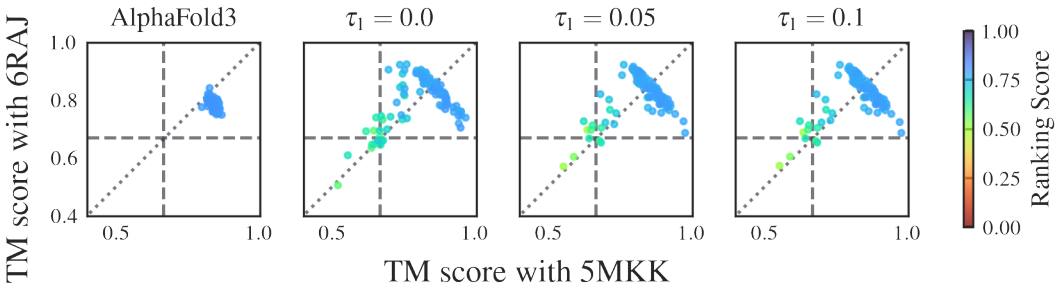
(g) Diversity plots of SLC1A1 for different τ_1 values.



(h) Diversity plots of SPF1 for different τ_1 values.

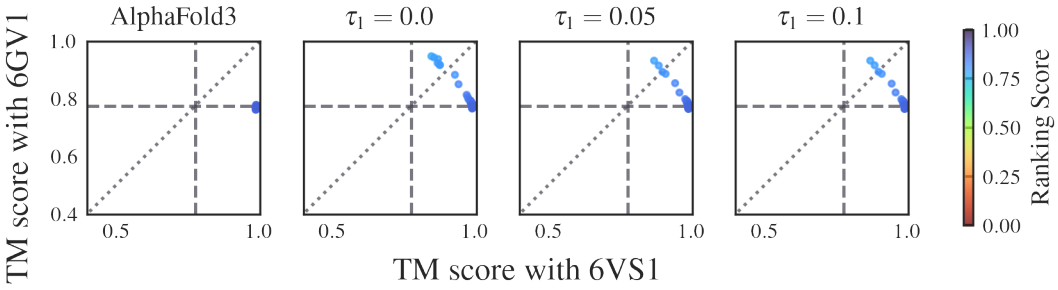


(i) Diversity plots of TM_0287 for different τ_1 values.

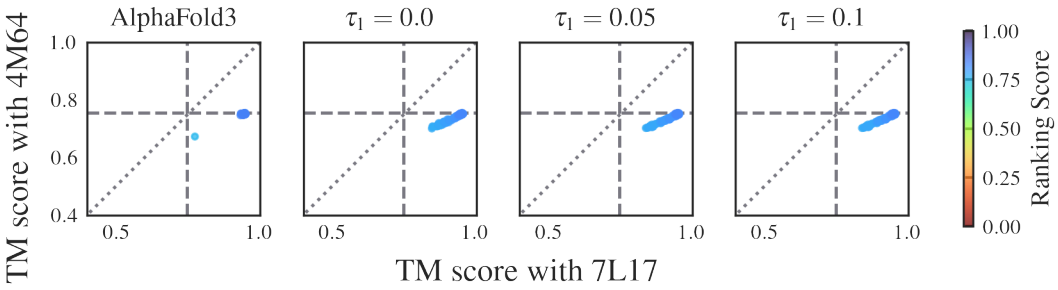


(j) Diversity plots of TT_C0976 for different τ_1 values.

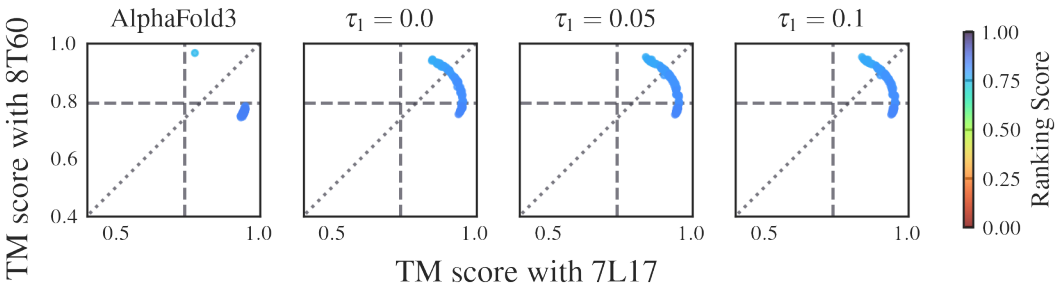
Figure 11: Effect of the CADS parameter τ_1 on the Transporters dataset (continued on next page).



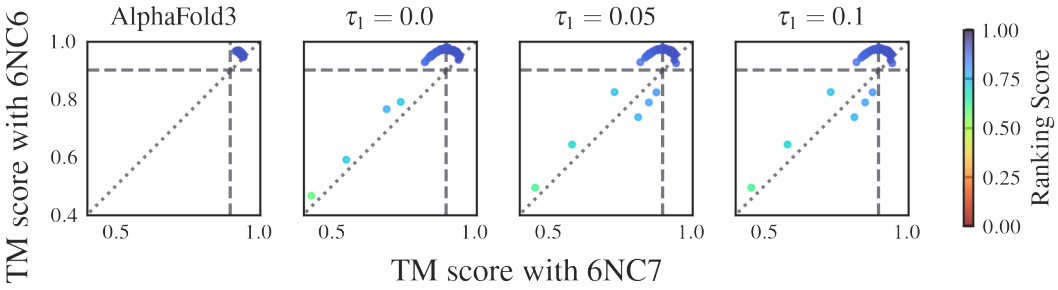
(k) Diversity plots of mdfA for different τ_1 values.



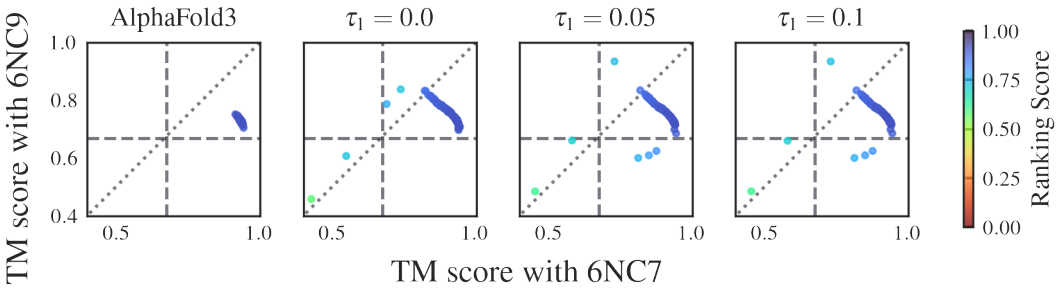
(l) Diversity plots of melB for different τ_1 values.



(m) Diversity plots of melB for different τ_1 values.

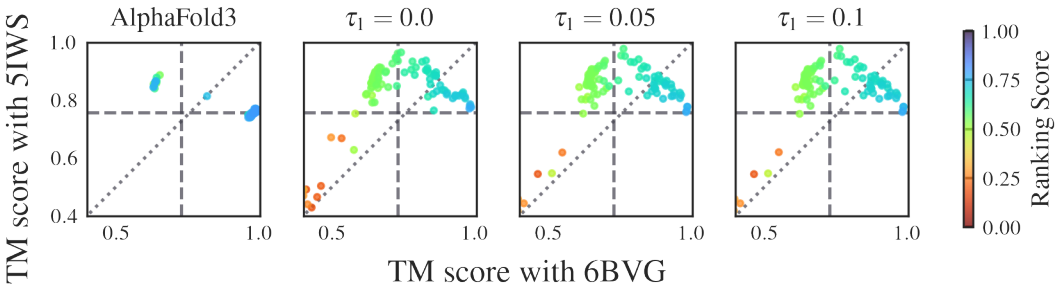


(n) Diversity plots of murJ for different τ_1 values.

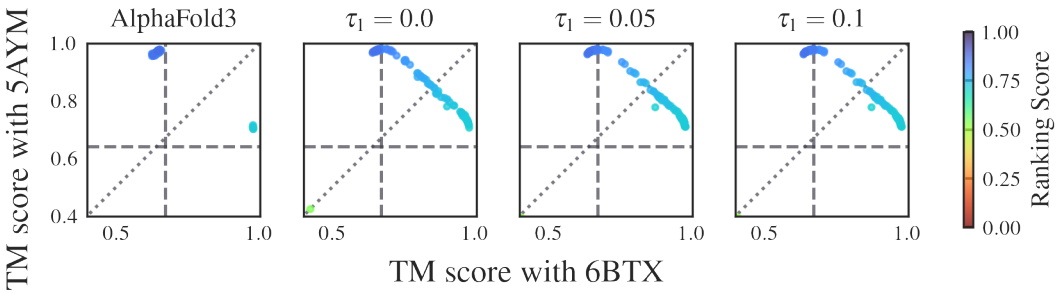


(o) Diversity plots of murJ for different τ_1 values.

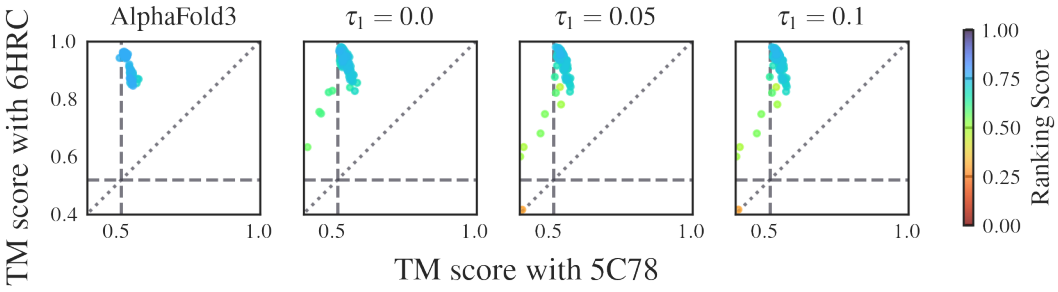
Figure 11: Effect of the CADS parameter τ_1 on the Transporters dataset (continued on next page).



(p) Diversity plots of ptsG for different τ_1 values.



(q) Diversity plots of slc39 for different τ_1 values.

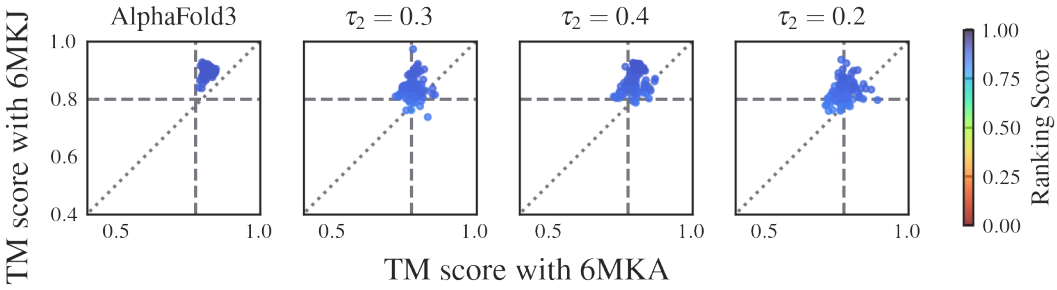


(r) Diversity plots of wlaB for different τ_1 values.

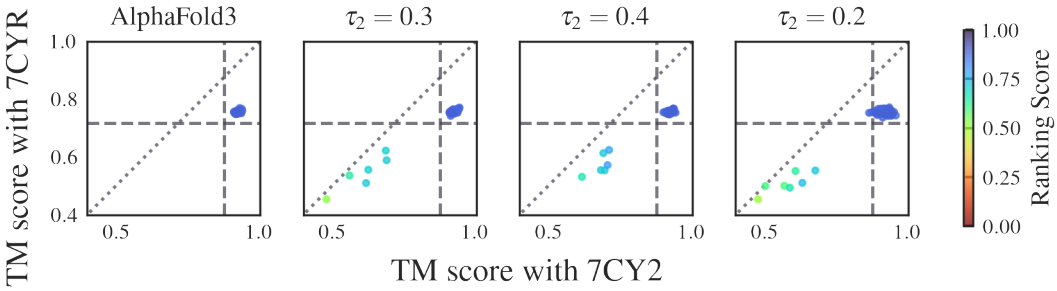
Figure 11: Effect of the CADS parameter τ_1 on the Transporters dataset.

D.4 VARYING THE VALUE OF τ_2

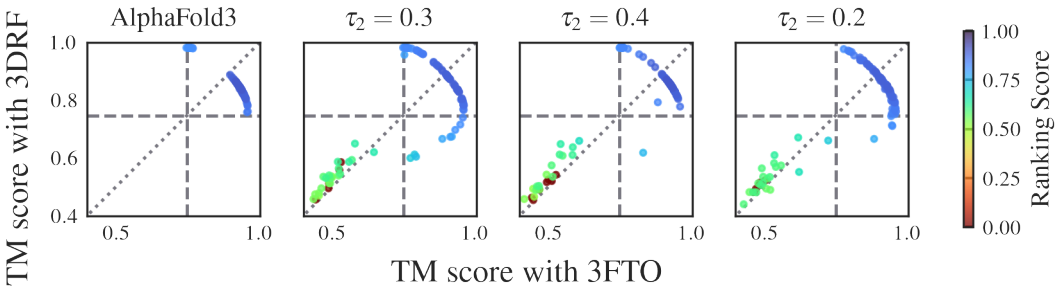
The value of τ_2 represents the number of diffusion condition corruption steps that will be performed with full strength as $\gamma(t) = 0$ for $t > \tau_2$. A smaller value of τ_2 cause to the corruption of the diffusion signal for a longer time frame. By default, predictions performed with AF3 GNMCADS uses $\tau_2 = 0.3$



(a) Diversity plots of A0A075Q0W3 for different τ_2 values.

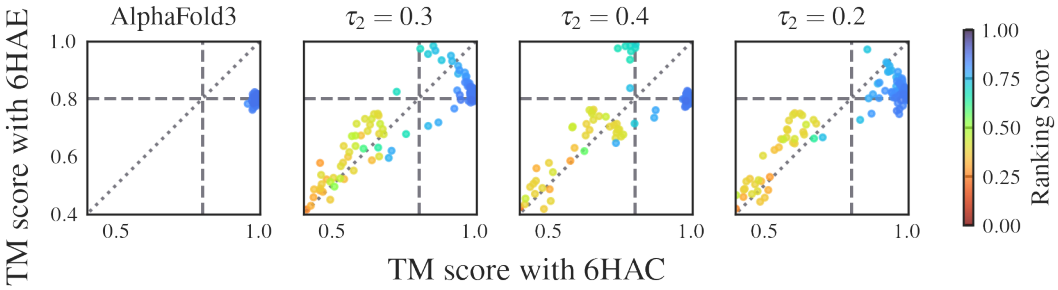


(b) Diversity plots of A0QTT2 for different τ_2 values.

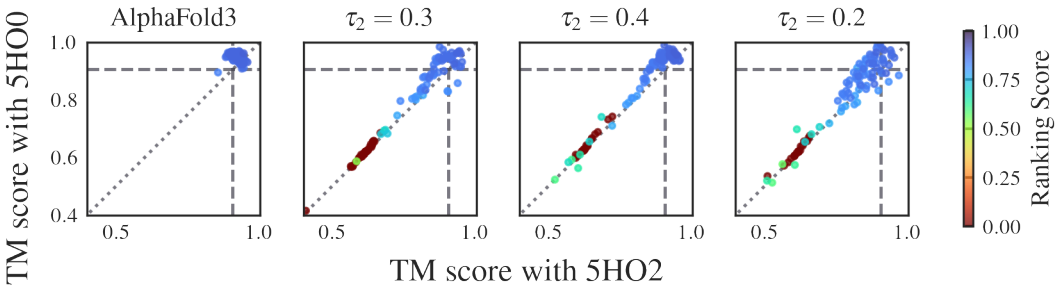


(c) Diversity plots of A2RJ53 for different τ_2 values.

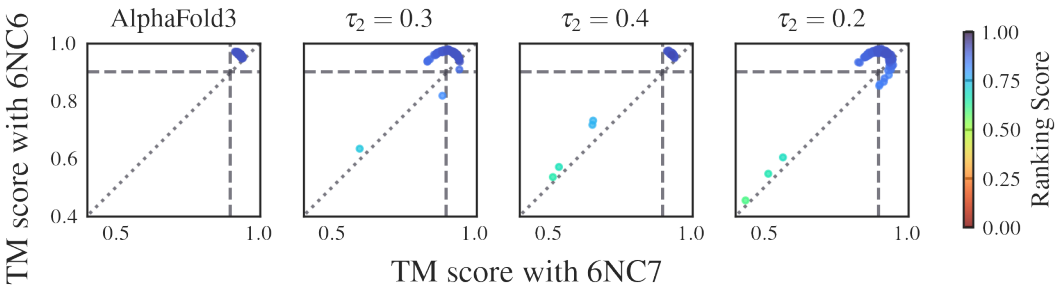
Figure 12: Effect of the CADS parameter τ_2 on the Open-Closed dataset (continued on next page).



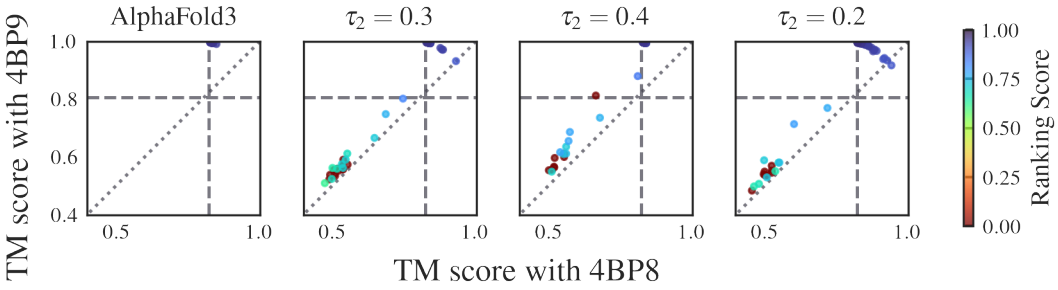
(d) Diversity plots of A6UVT1 for different τ_2 values.



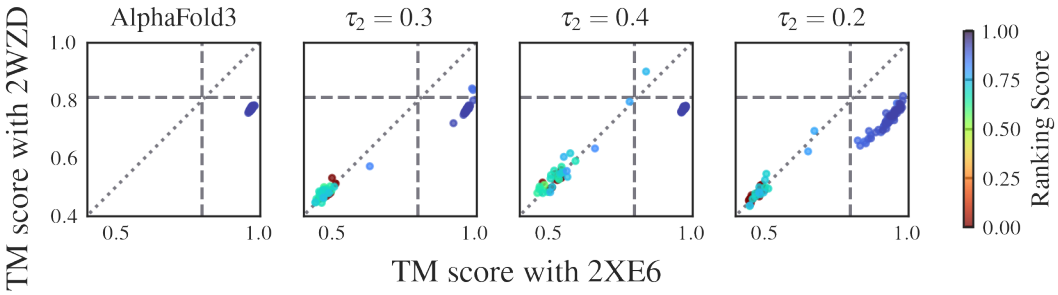
(e) Diversity plots of B3EYN2 for different τ_2 values.



(f) Diversity plots of B7IE18 for different τ_2 values.

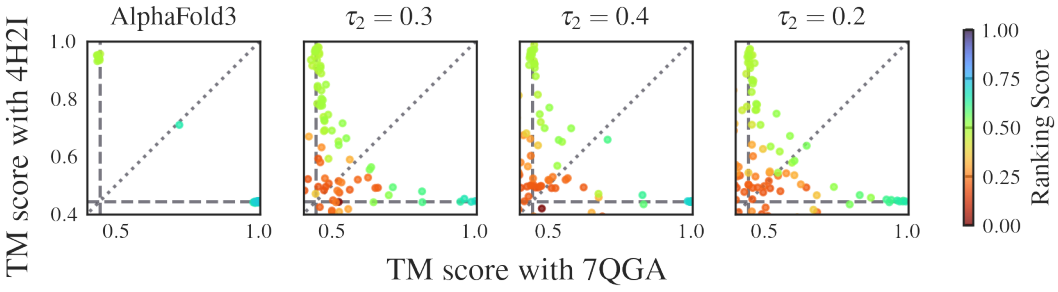


(g) Diversity plots of O76728 for different τ_2 values.

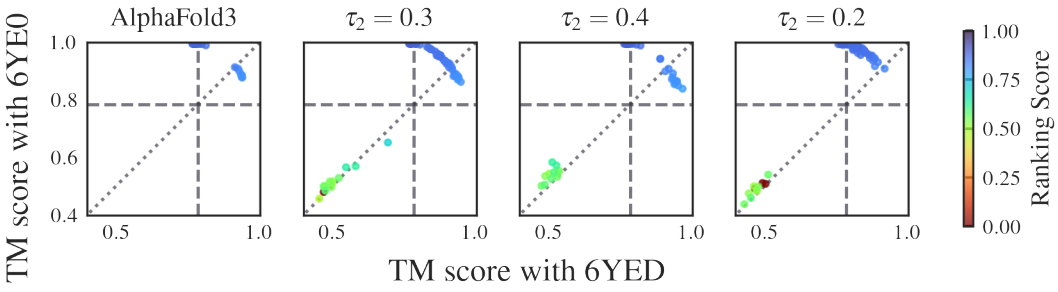


(h) Diversity plots of P00558 for different τ_2 values.

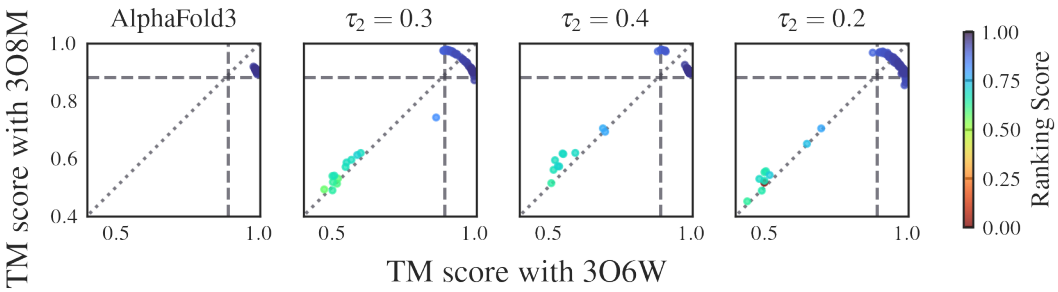
Figure 12: Effect of the CADS parameter τ_2 on the Open-Closed dataset (continued on next page).



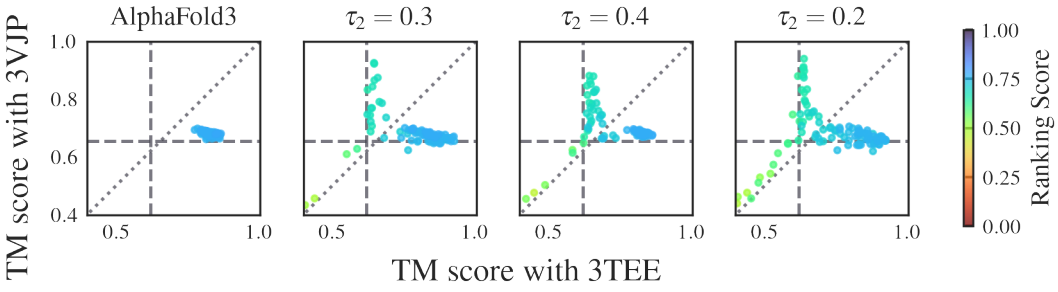
(i) Diversity plots of P21589 for different τ_2 values.



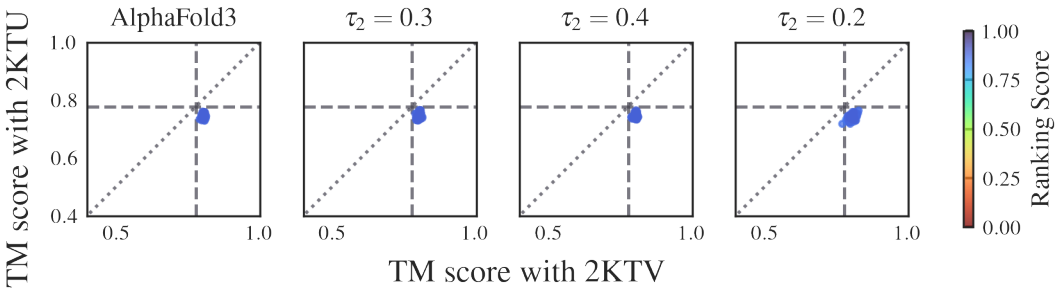
(j) Diversity plots of P31133 for different τ_2 values.



(k) Diversity plots of P33284 for different τ_2 values.

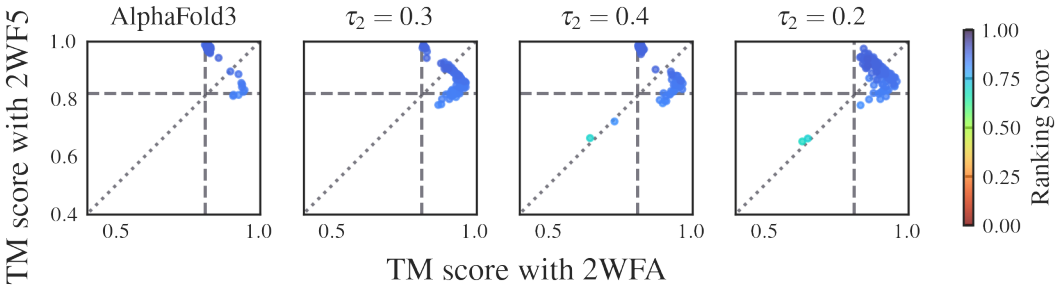


(l) Diversity plots of P40131 for different τ_2 values.

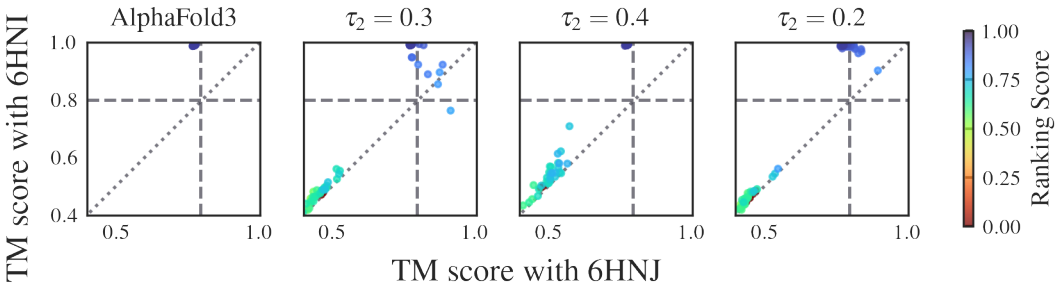


(m) Diversity plots of P62495 for different τ_2 values.

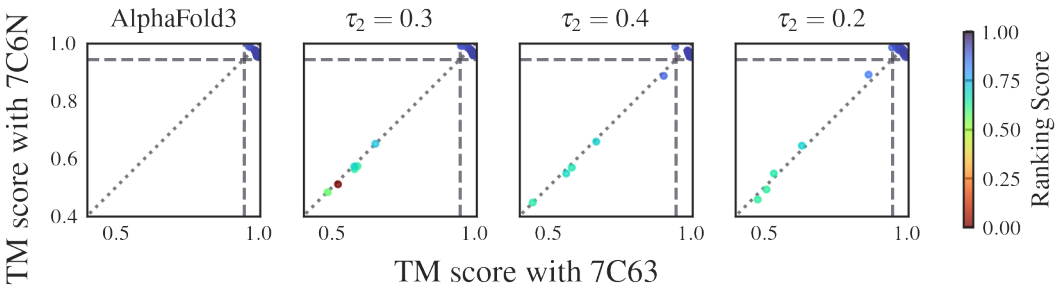
Figure 12: Effect of the CADS parameter τ_2 on the Open-Closed dataset (continued on next page).



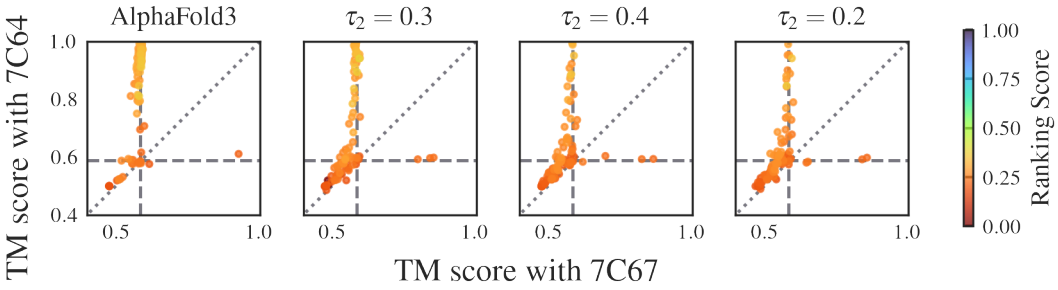
(n) Diversity plots of P71447 for different τ_2 values.



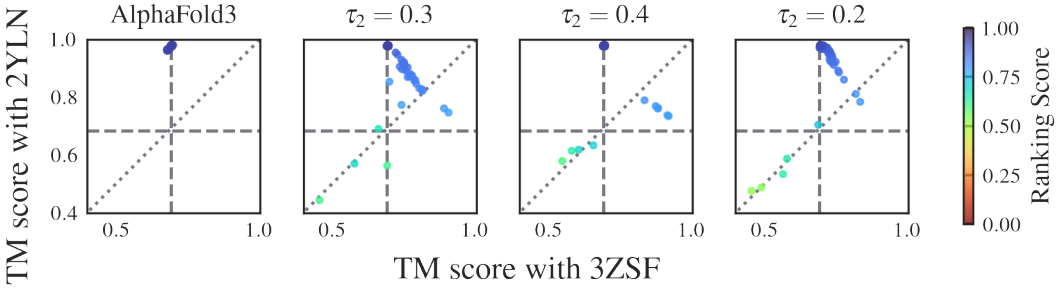
(o) Diversity plots of Q18A65 for different τ_2 values.



(p) Diversity plots of Q53W80 for different τ_2 values.

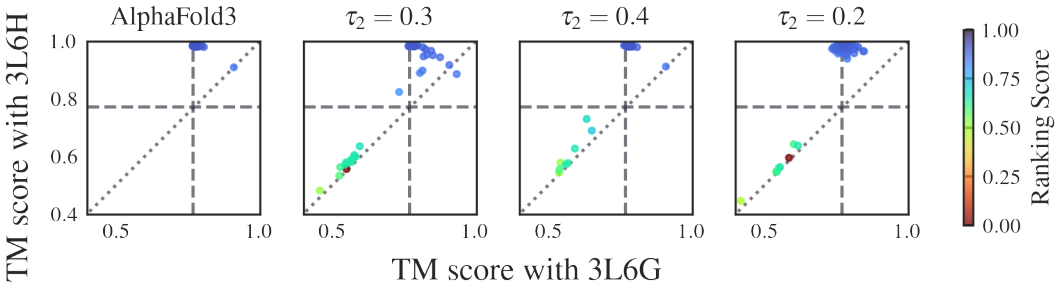


(q) Diversity plots of Q53W80 for different τ_2 values.

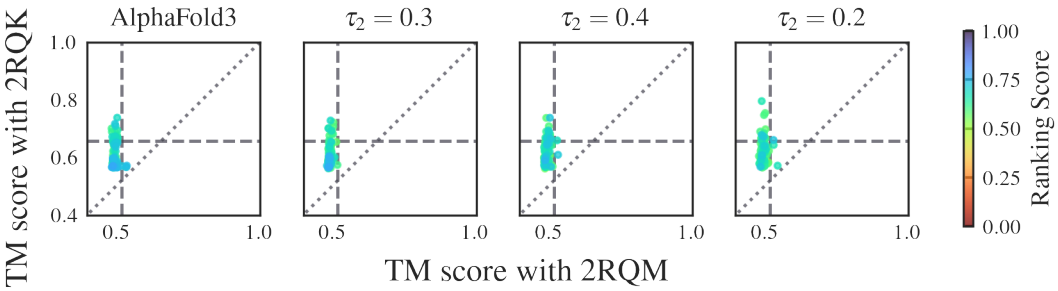


(r) Diversity plots of Q5F9M1 for different τ_2 values.

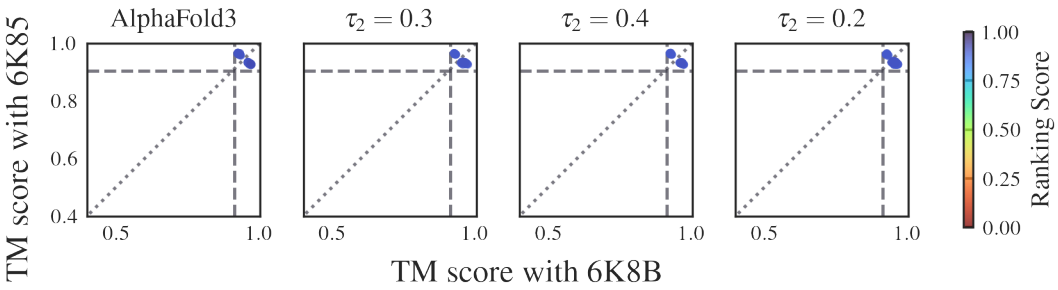
Figure 12: Effect of the CADS parameter τ_2 on the Open-Closed dataset (continued on next page).



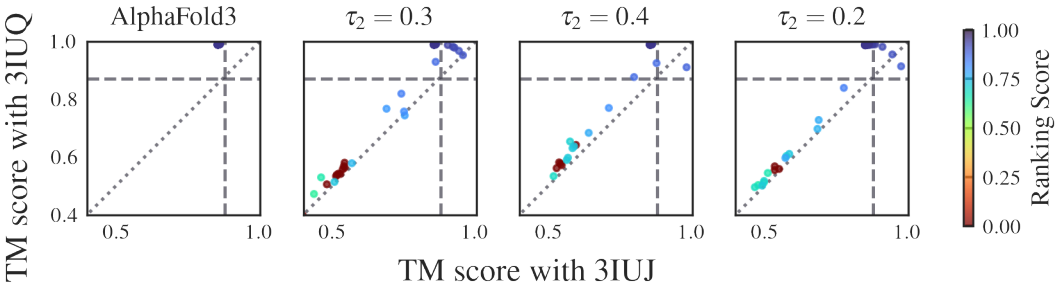
(s) Diversity plots of Q7DAU8 for different τ_2 values.



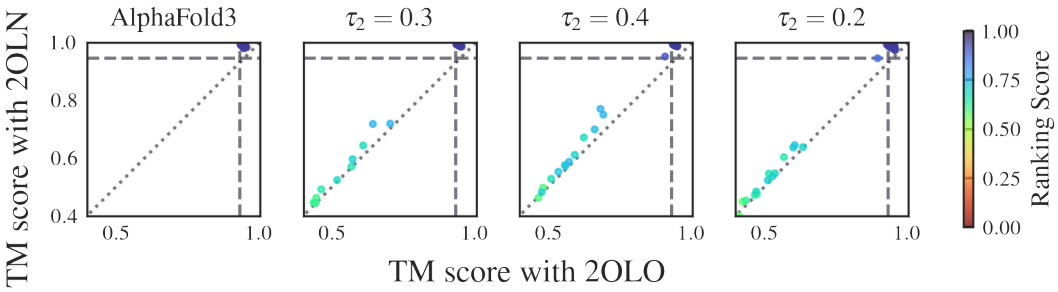
(t) Diversity plots of Q9ERE7 for different τ_2 values.



(u) Diversity plots of Q9SS90 for different τ_2 values.

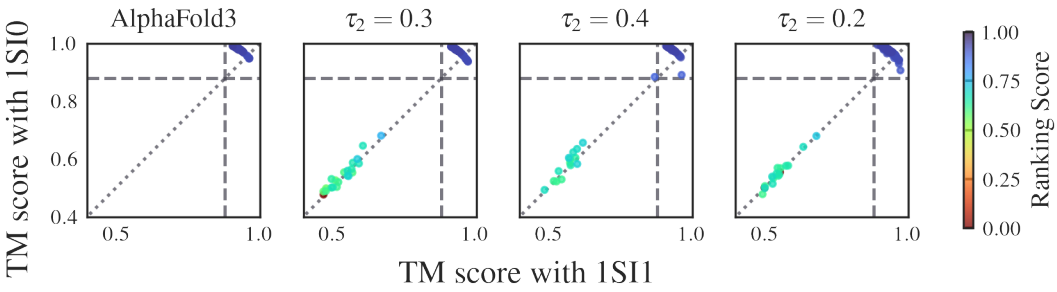


(v) Diversity plots of Q9X6R4 for different τ_2 values.



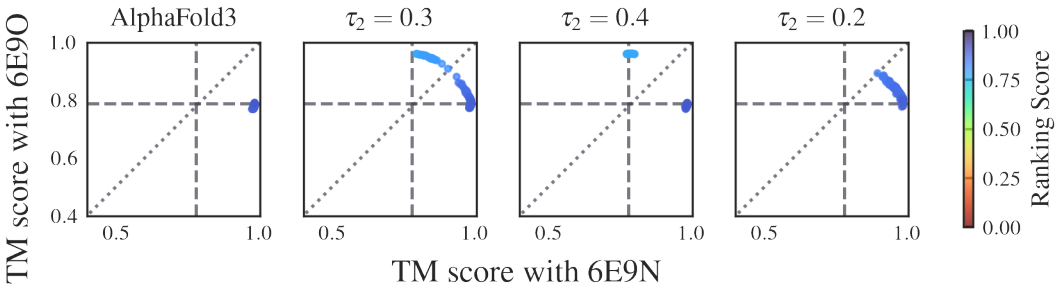
(w) Diversity plots of Q9X9P9 for different τ_2 values.

Figure 12: Effect of the CADS parameter τ_2 on the Open-Closed dataset (continued on next page).

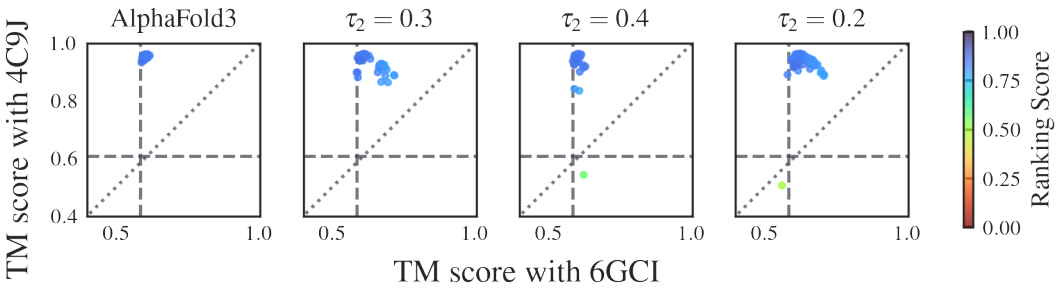


(x) Diversity plots of Q9Z4N6 for different τ_2 values.

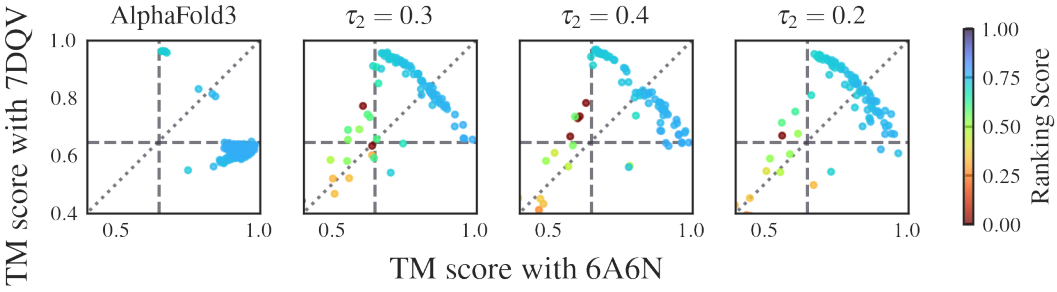
Figure 12: Effect of the CADS parameter τ_2 on the Open-Closed dataset.



(a) Diversity plots of A5U30_003247 for different τ_2 values.

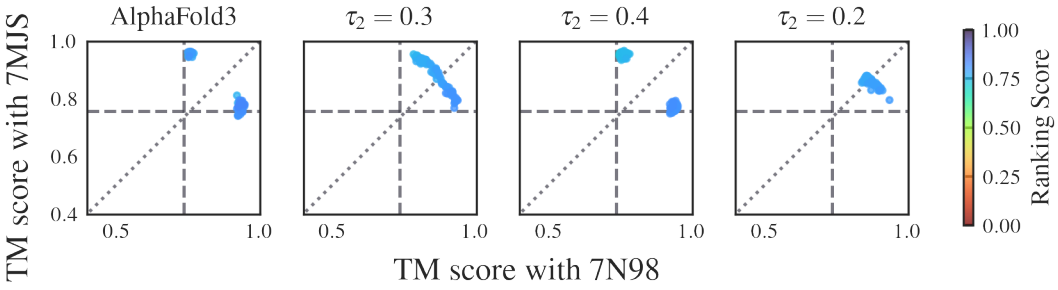


(b) Diversity plots of AAC3 for different τ_2 values.

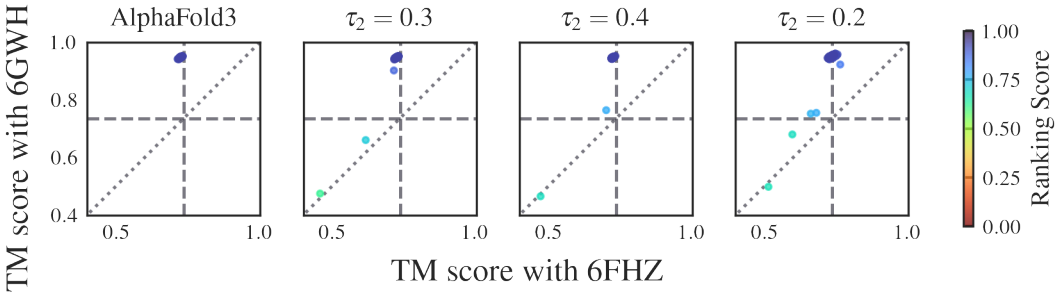


(c) Diversity plots of CYME_CMD148C for different τ_2 values.

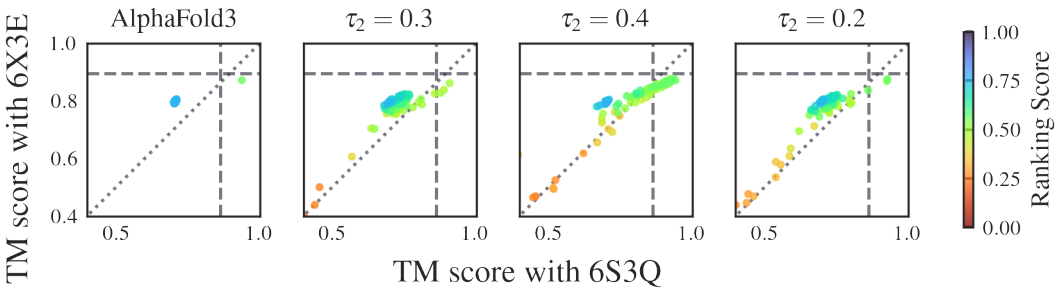
Figure 13: Effect of the CADS parameter τ_2 on the Transporters dataset (continued on next page).



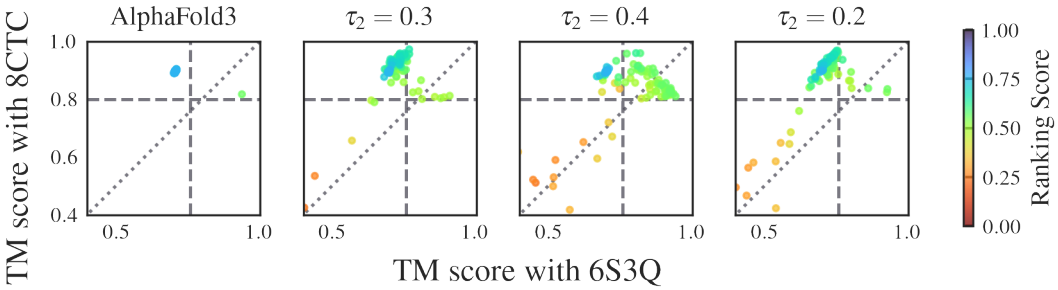
(d) Diversity plots of MFSD2A for different τ_2 values.



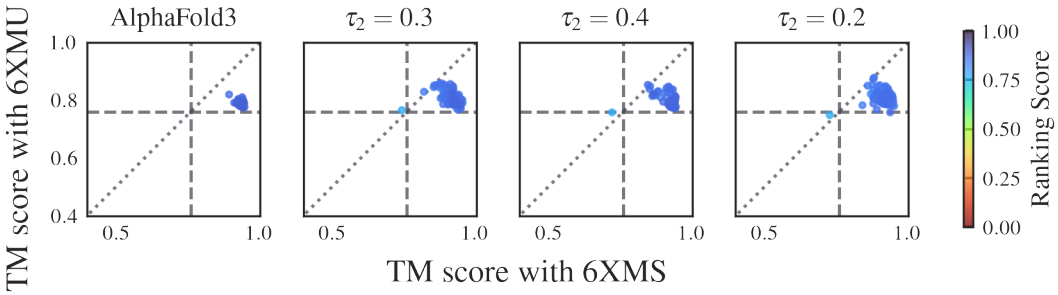
(e) Diversity plots of PF0708 for different τ_2 values.



(f) Diversity plots of SLC1A1 for different τ_2 values.

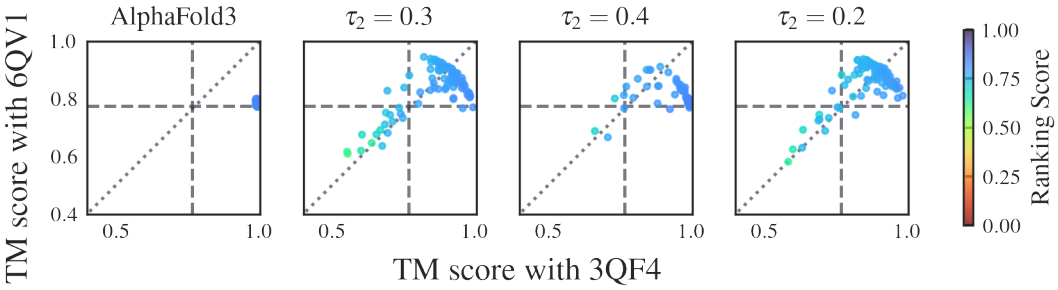


(g) Diversity plots of SLC1A1 for different τ_2 values.

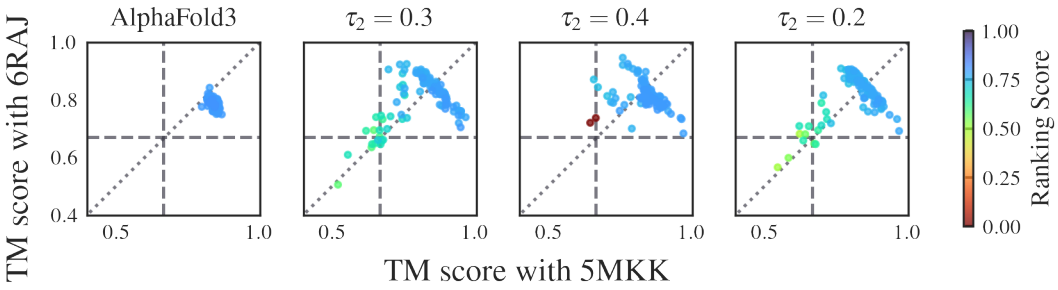


(h) Diversity plots of SPF1 for different τ_2 values.

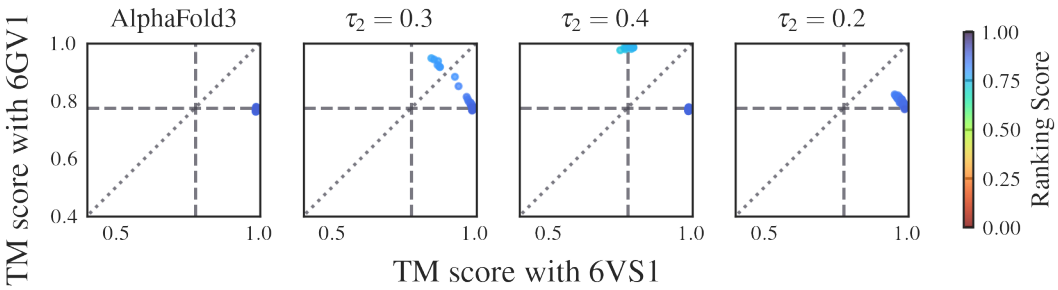
Figure 13: Effect of the CADS parameter τ_2 on the Transporters dataset (continued on next page).



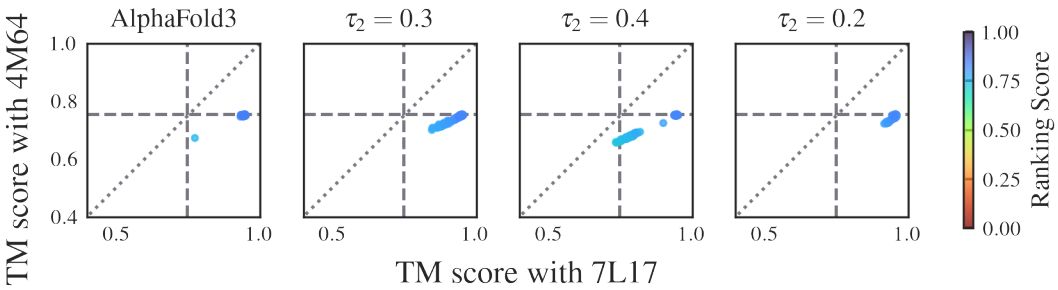
(i) Diversity plots of TM_0287 for different τ_2 values.



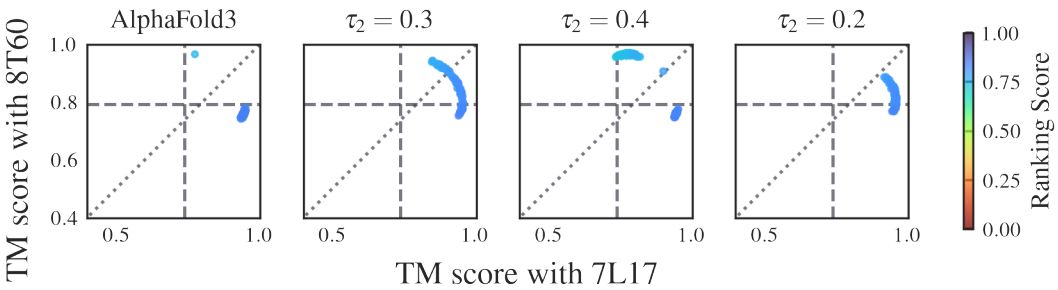
(j) Diversity plots of TT_C0976 for different τ_2 values.



(k) Diversity plots of mdfA for different τ_2 values.

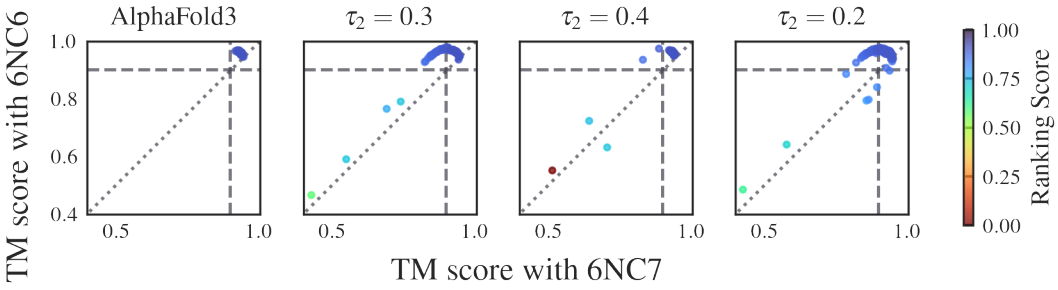


(l) Diversity plots of melB for different τ_2 values.

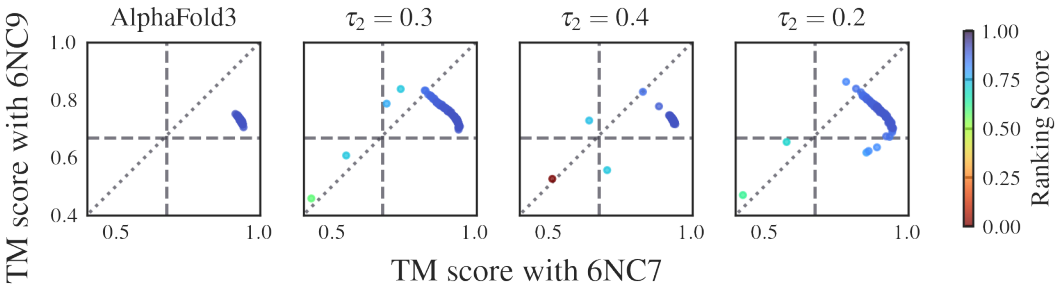


(m) Diversity plots of melB for different τ_2 values.

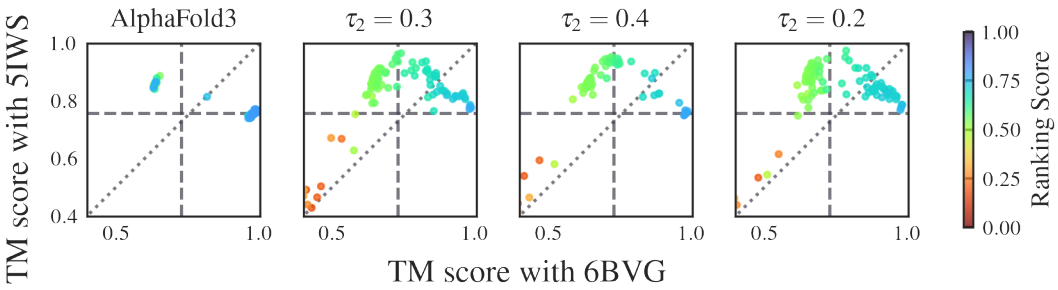
Figure 13: Effect of the CADS parameter τ_2 on the Transporters dataset (continued on next page).



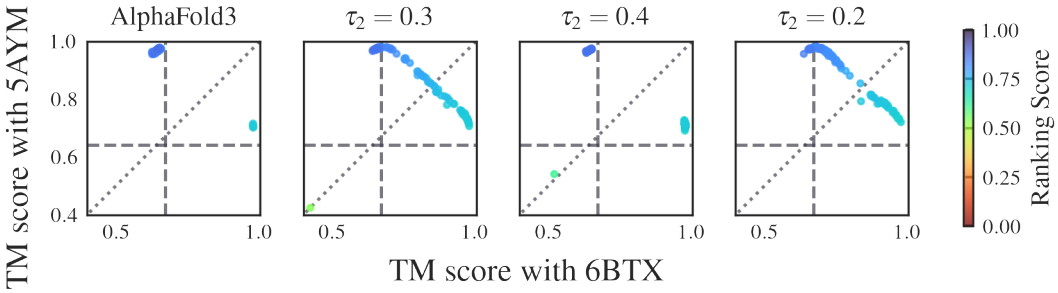
(n) Diversity plots of murJ for different τ_2 values.



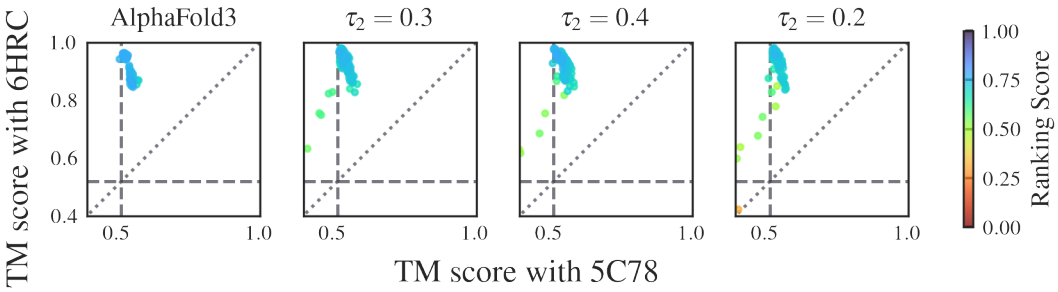
(o) Diversity plots of murJ for different τ_2 values.



(p) Diversity plots of ptsG for different τ_2 values.



(q) Diversity plots of slc39 for different τ_2 values.

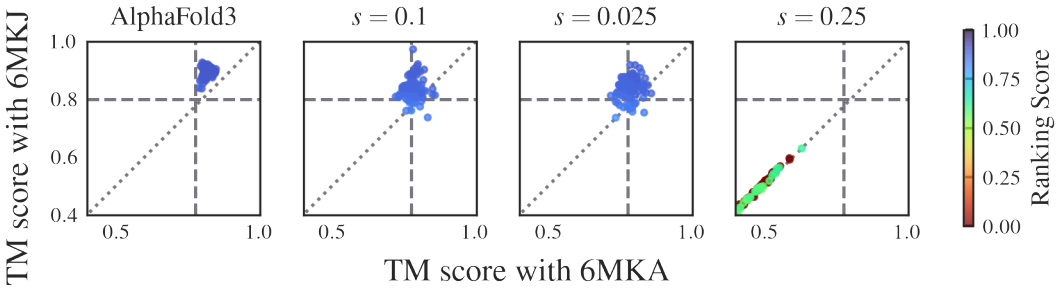


(r) Diversity plots of wlaB for different τ_2 values.

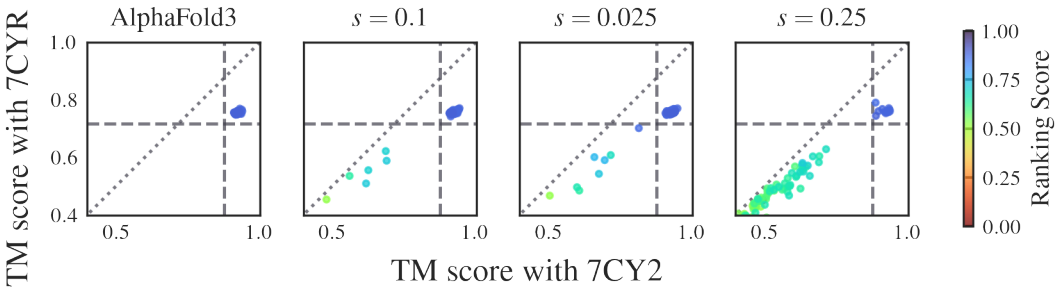
Figure 13: Effect of the CADS parameter τ_2 on the Transporters dataset.

D.5 VARYING THE VALUE OF NOISE SCALE

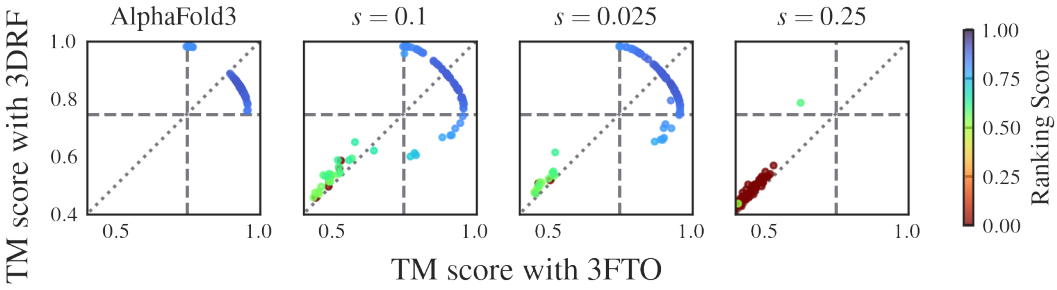
The value of s represents the noise scale of condition annealing. In AF3 GNMCADS, the noise scale of $s = 0.1$ is used by default.



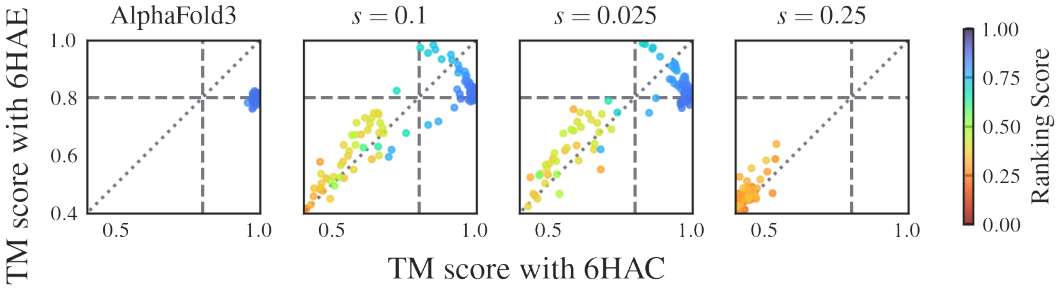
(a) Diversity plots of A0A075Q0W3 for different noise scale (s) values.



(b) Diversity plots of A0QTT2 for different noise scale (s) values.

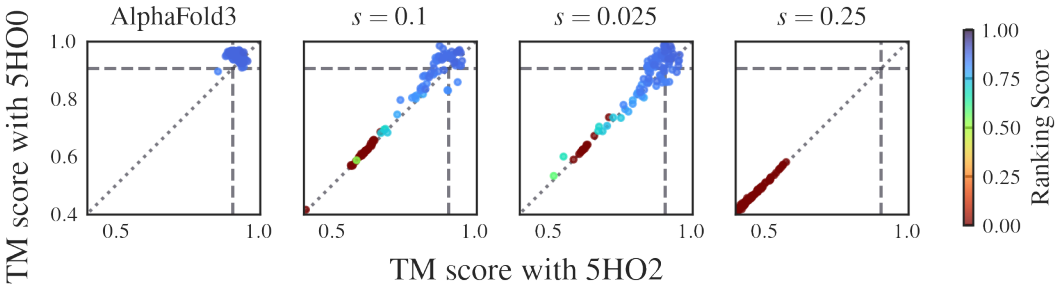


(c) Diversity plots of A2RJ53 for different noise scale (s) values.

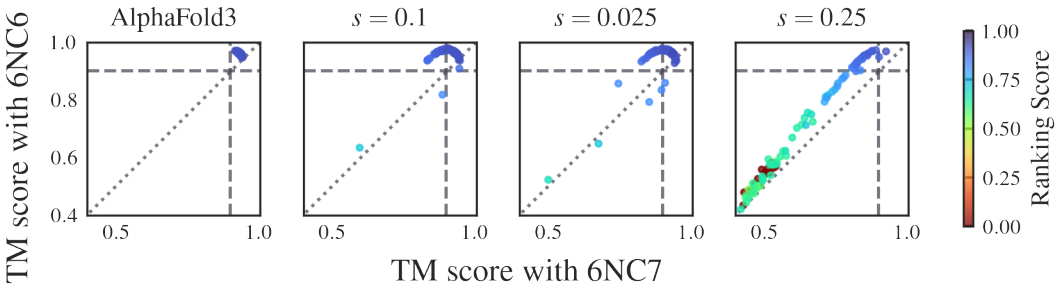


(d) Diversity plots of A6UVT1 for different noise scale (s) values.

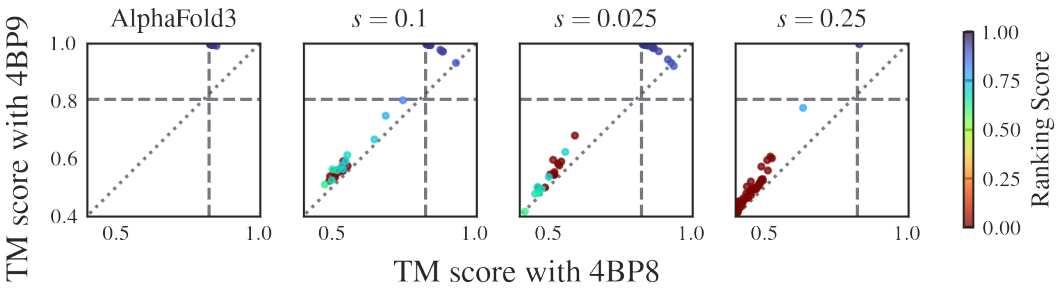
Figure 14: Effect of the CADS parameter noise scale (s) on the Open-Closed dataset (continued on next page).



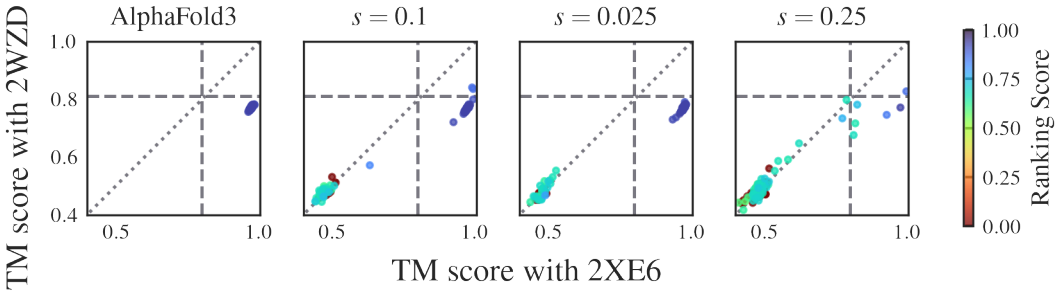
(e) Diversity plots of B3EYN2 for different noise scale (s) values.



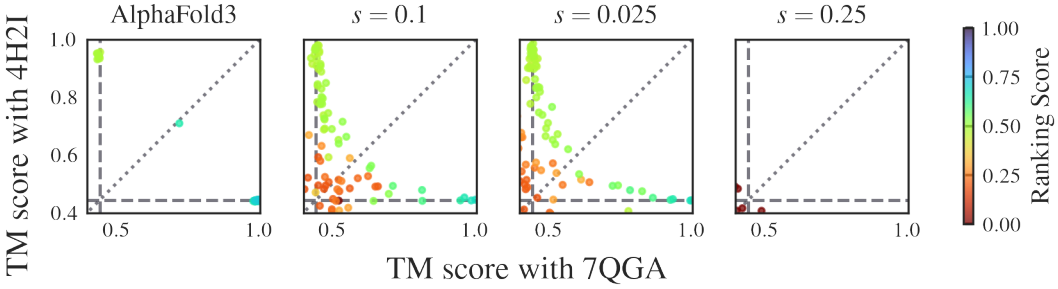
(f) Diversity plots of B7IE18 for different noise scale (s) values.



(g) Diversity plots of O76728 for different noise scale (s) values.

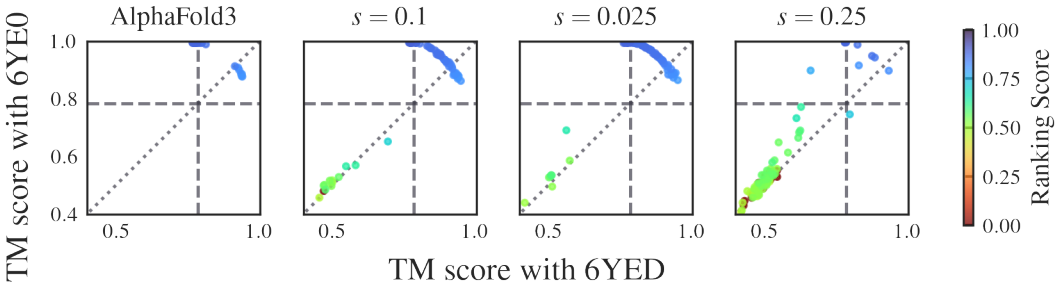


(h) Diversity plots of P00558 for different noise scale (s) values.

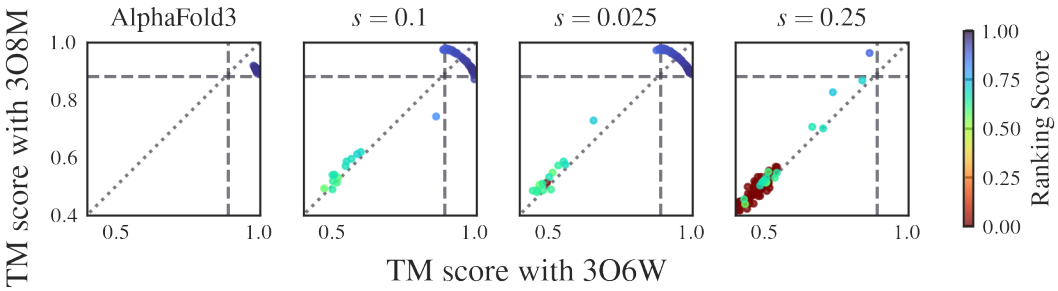


(i) Diversity plots of P21589 for different noise scale (s) values.

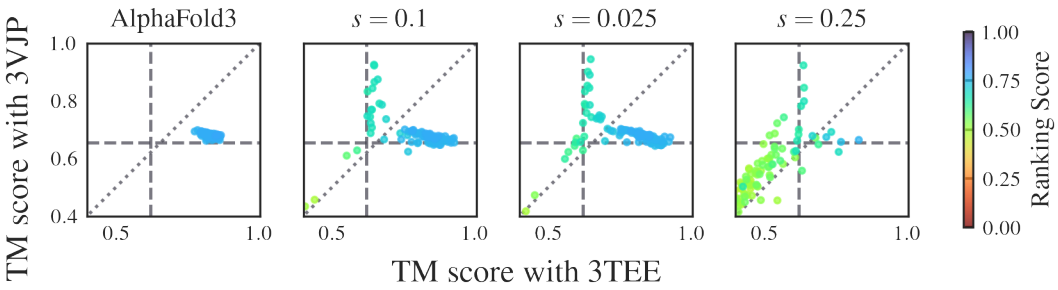
Figure 14: Effect of the CADS parameter noise scale (s) on the Open-Closed dataset (continued on next page).



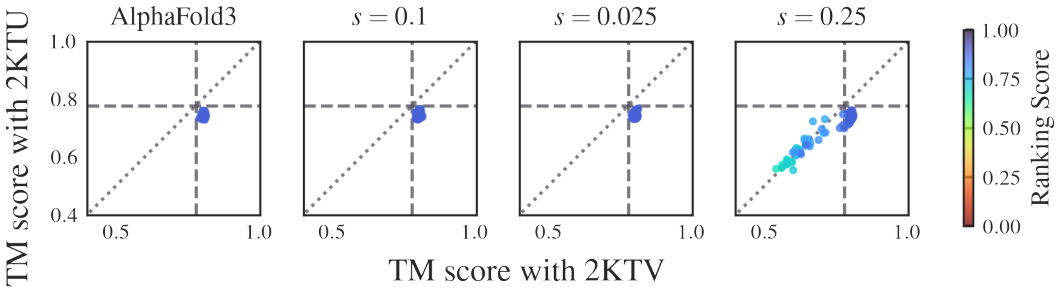
(j) Diversity plots of P31133 for different noise scale (s) values.



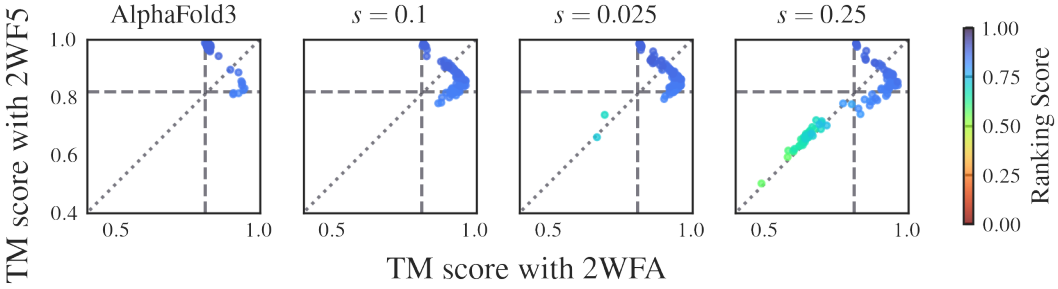
(k) Diversity plots of P33284 for different noise scale (s) values.



(l) Diversity plots of P40131 for different noise scale (s) values.

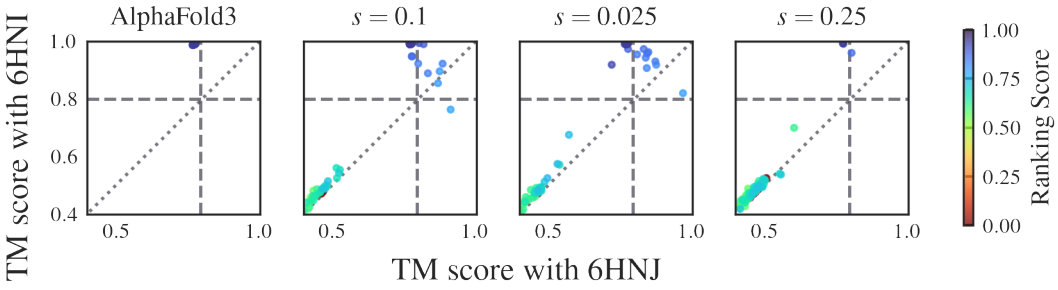


(m) Diversity plots of P62495 for different noise scale (s) values.

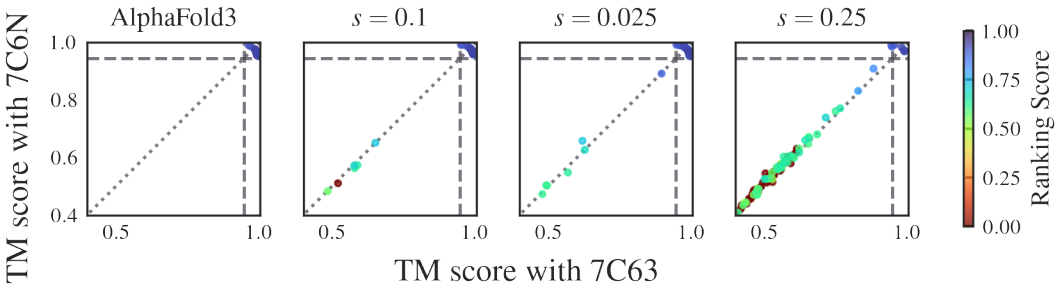


(n) Diversity plots of P71447 for different noise scale (s) values.

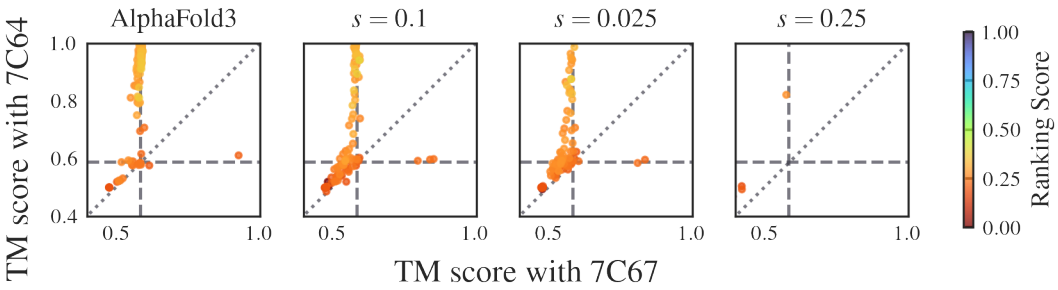
Figure 14: Effect of the CADS parameter noise scale (s) on the Open-Closed dataset (continued on next page).



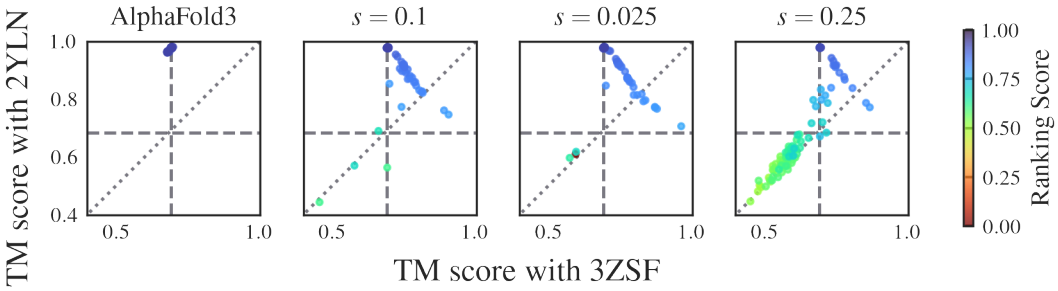
(o) Diversity plots of Q18A65 for different noise scale (s) values.



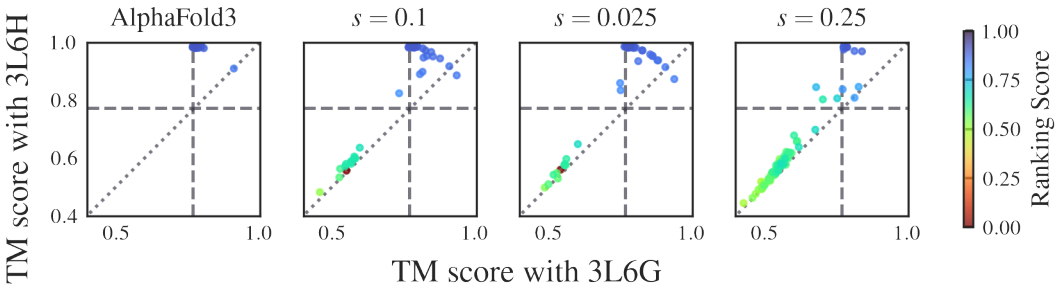
(p) Diversity plots of Q53W80 for different noise scale (s) values.



(q) Diversity plots of Q53W80 for different noise scale (s) values.

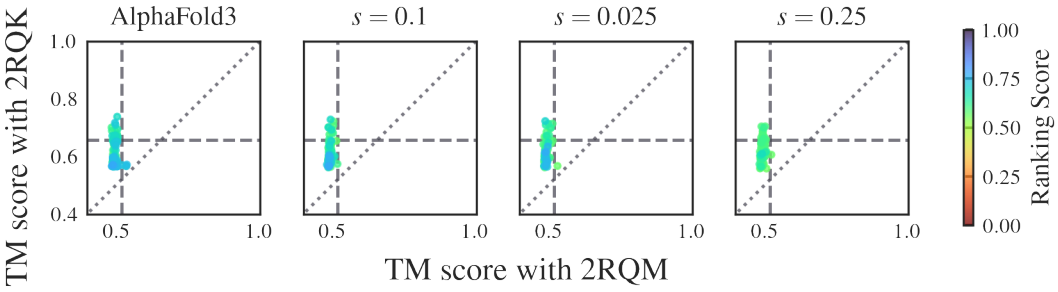


(r) Diversity plots of Q5F9M1 for different noise scale (s) values.

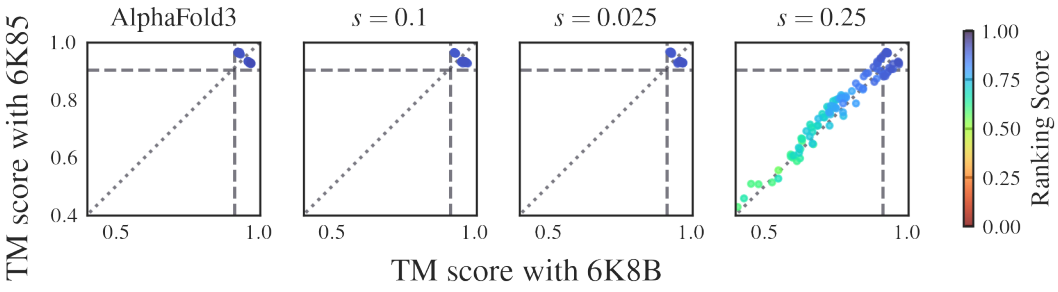


(s) Diversity plots of Q7DAU8 for different noise scale (s) values.

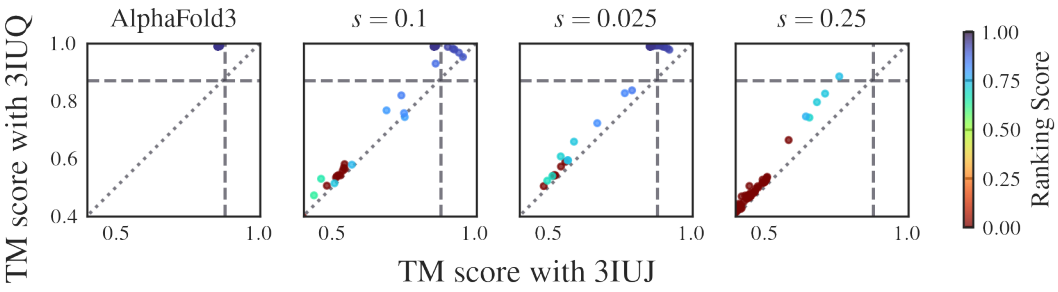
Figure 14: Effect of the CADS parameter noise scale (s) on the Open-Closed dataset (continued on next page).



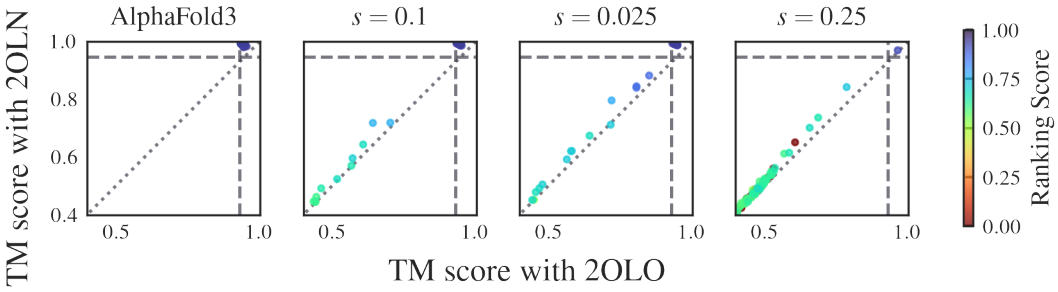
(t) Diversity plots of Q9ERE7 for different noise scale (s) values.



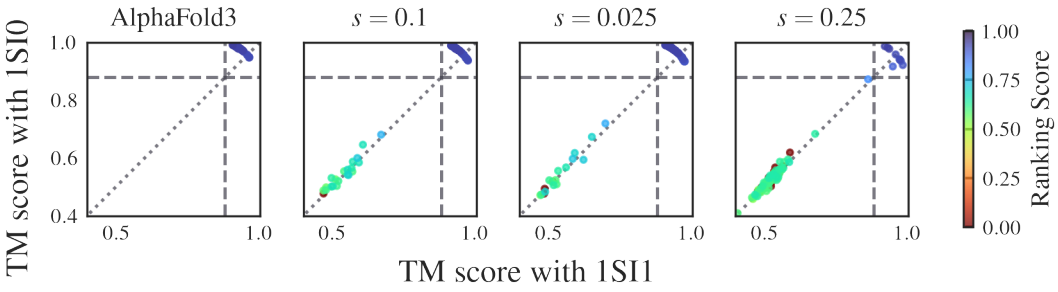
(u) Diversity plots of Q9SS90 for different noise scale (s) values.



(v) Diversity plots of Q9X6R4 for different noise scale (s) values.

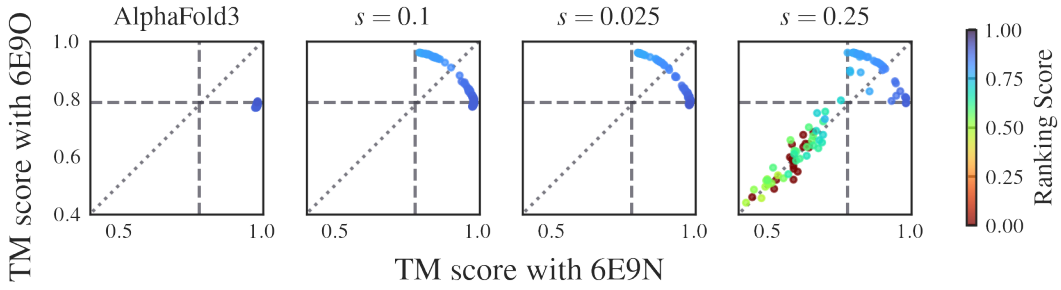


(w) Diversity plots of Q9X9P9 for different noise scale (s) values.

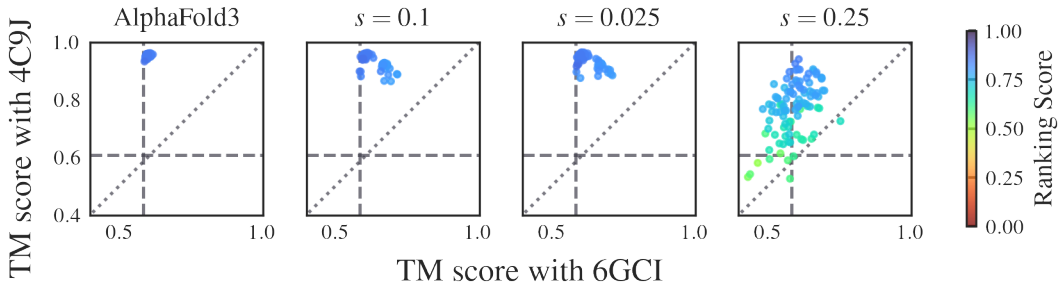


(x) Diversity plots of Q9Z4N6 for different noise scale (s) values.

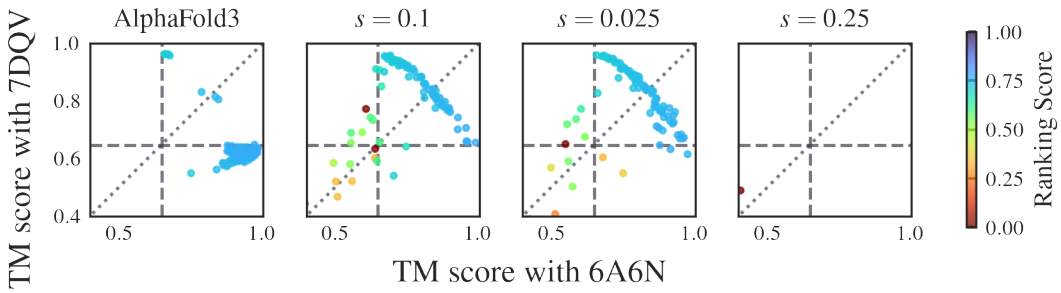
Figure 14: Effect of the CADS parameter noise scale (s) on the Open-Closed dataset.



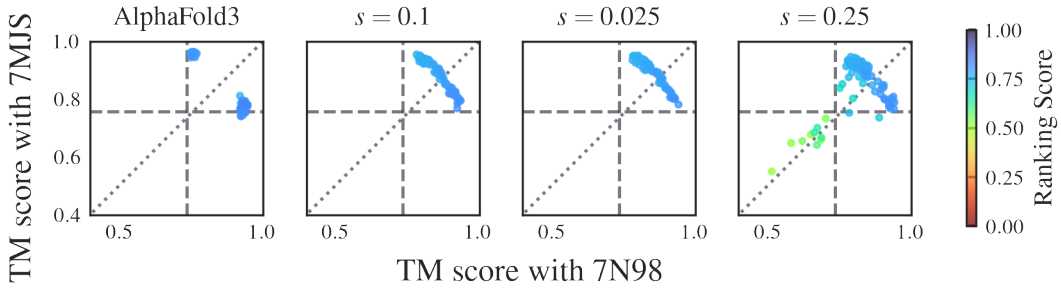
(a) Diversity plots of A5U30_003247 for different noise scale (s) values.



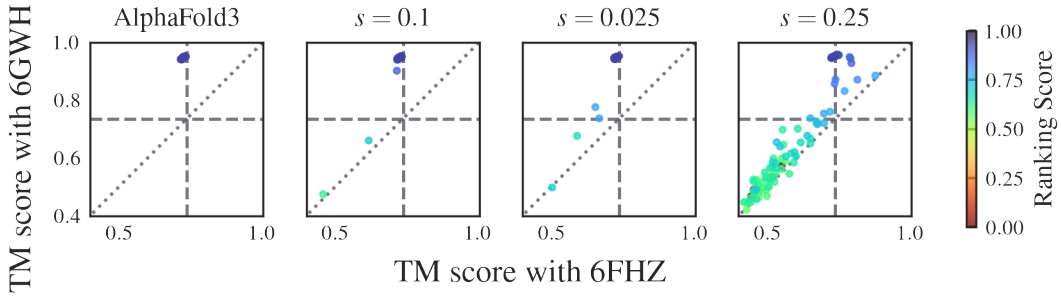
(b) Diversity plots of AAC3 for different noise scale (s) values.



(c) Diversity plots of CYME_CMD148C for different noise scale (s) values.

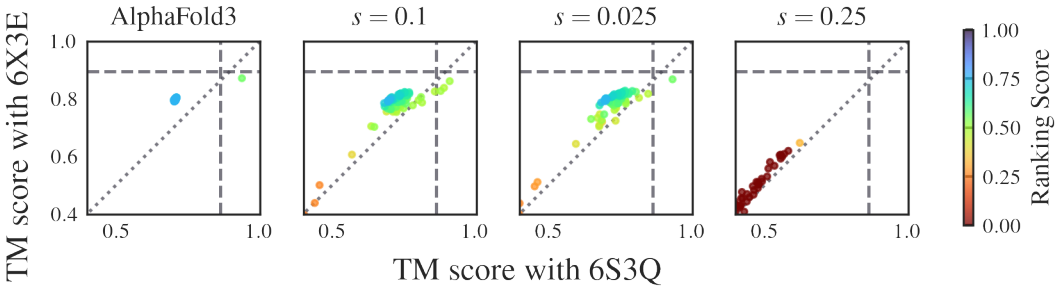


(d) Diversity plots of MFSD2A for different noise scale (s) values.

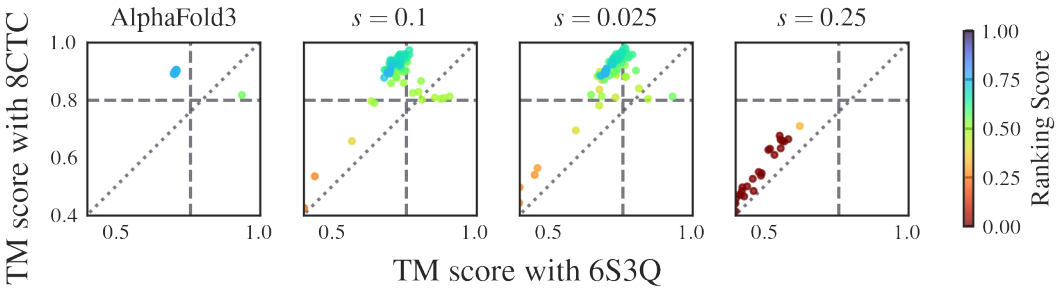


(e) Diversity plots of PF0708 for different noise scale (s) values.

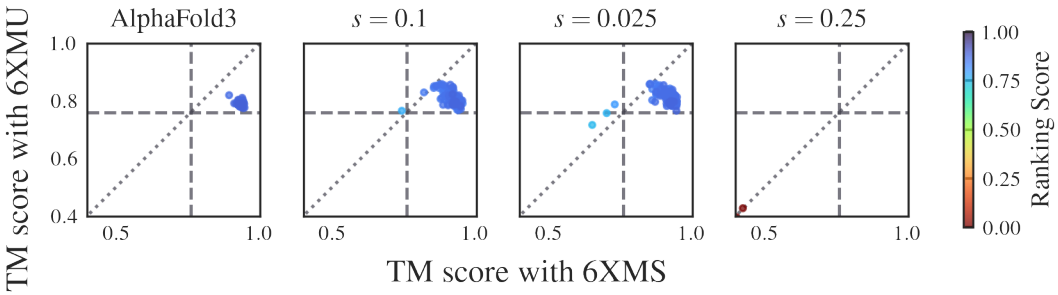
Figure 15: Effect of the CADS parameter noise scale (s) on the Transporters dataset (continued on next page).



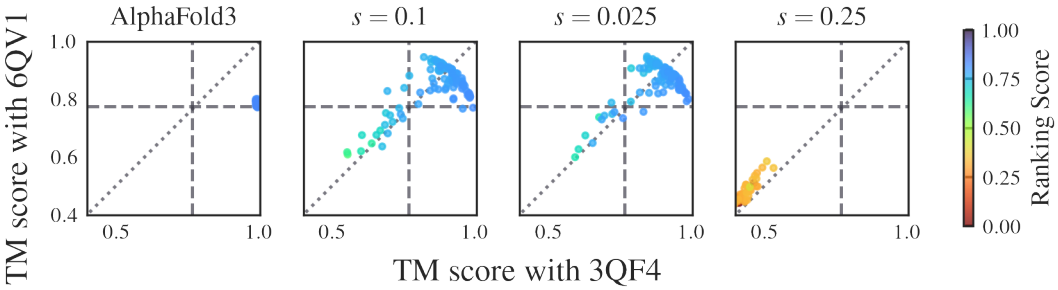
(f) Diversity plots of SLC1A1 for different noise scale (s) values.



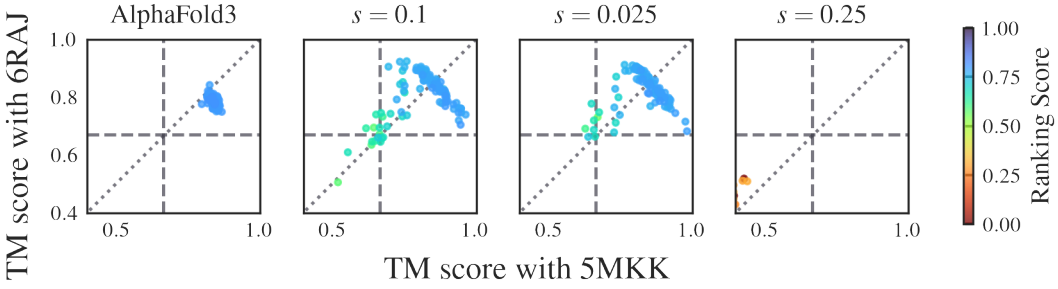
(g) Diversity plots of SLC1A1 for different noise scale (s) values.



(h) Diversity plots of SPF1 for different noise scale (s) values.

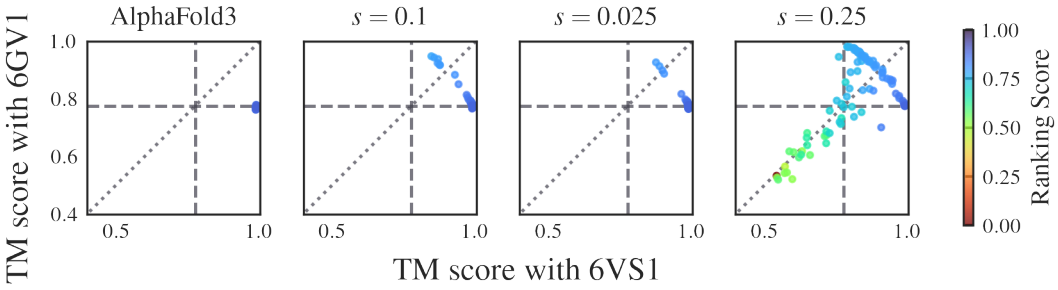


(i) Diversity plots of TM_0287 for different noise scale (s) values.

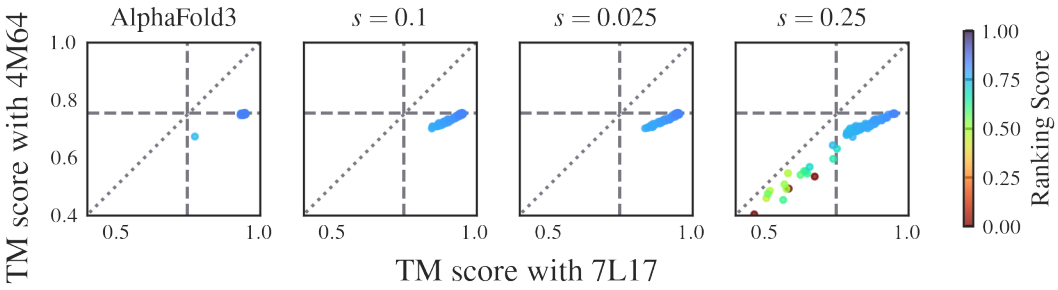


(j) Diversity plots of TT_C0976 for different noise scale (s) values.

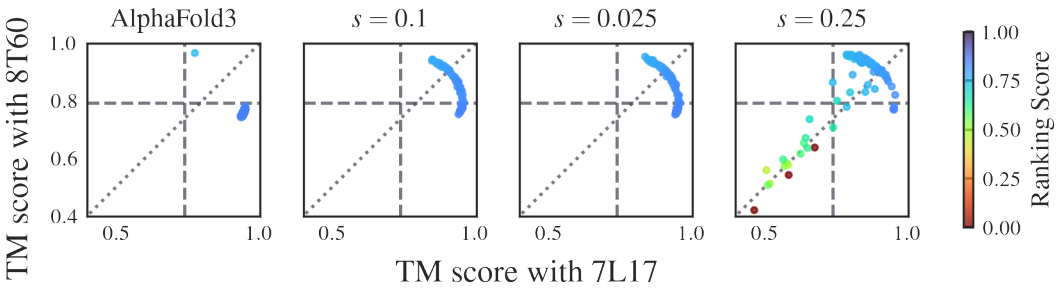
Figure 15: Effect of the CADS parameter noise scale (s) on the Transporters dataset (continued on next page).



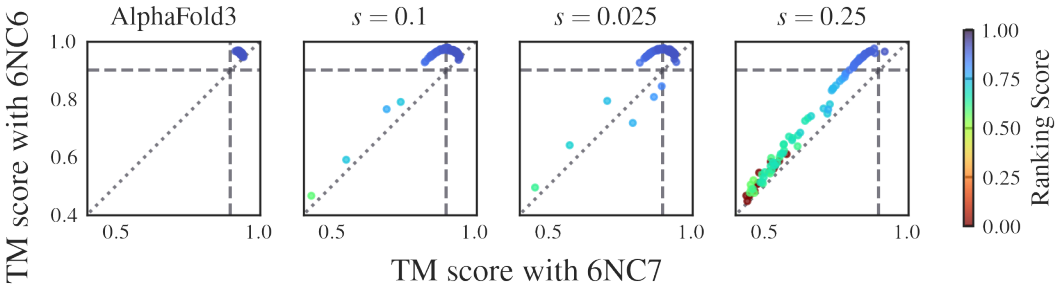
(k) Diversity plots of mdfA for different noise scale (s) values.



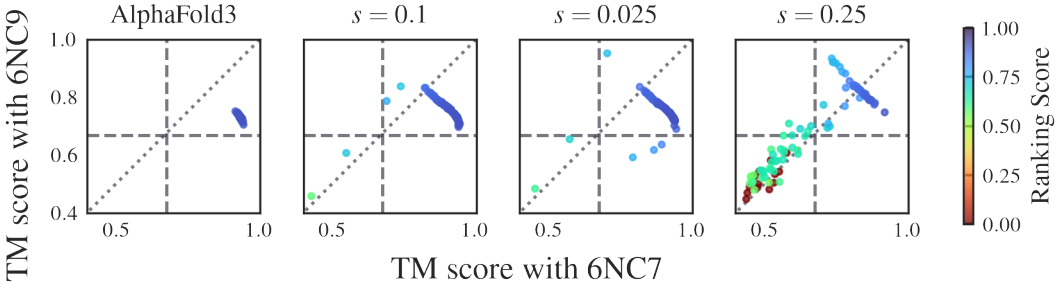
(l) Diversity plots of melB for different noise scale (s) values.



(m) Diversity plots of melB for different noise scale (s) values.

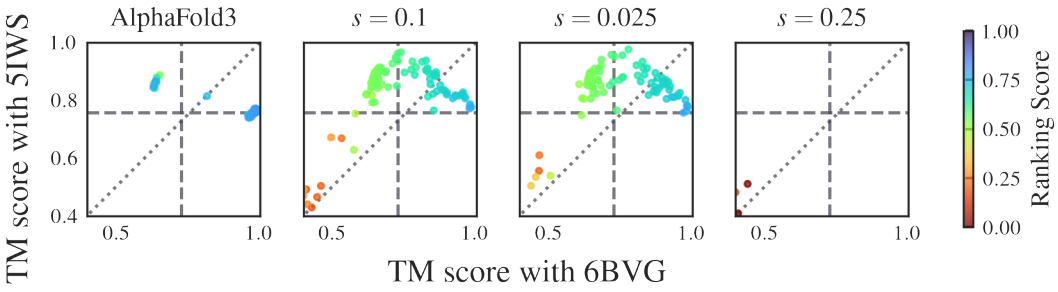


(n) Diversity plots of murJ for different noise scale (s) values.

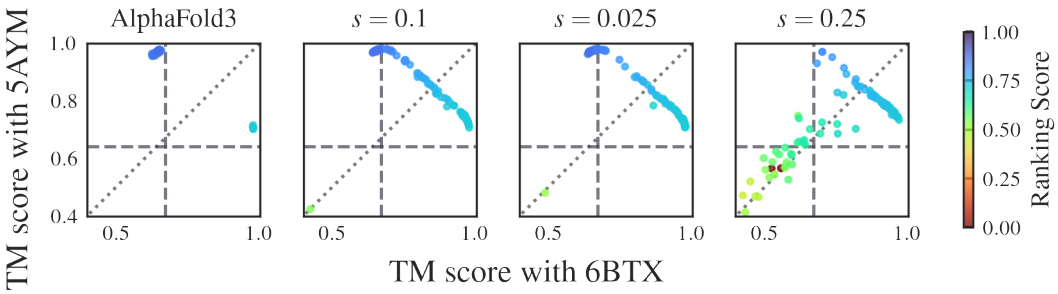


(o) Diversity plots of murJ for different noise scale (s) values.

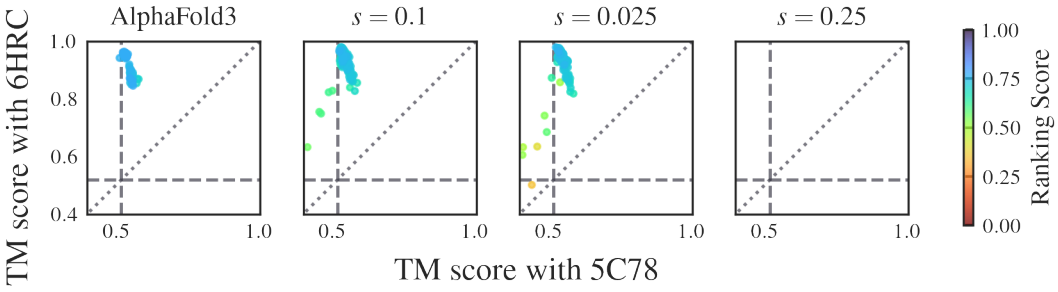
Figure 15: Effect of the CADS parameter noise scale (s) on the Transporters dataset (continued on next page).



(p) Diversity plots of ptsG for different noise scale (s) values.



(q) Diversity plots of slc39 for different noise scale (s) values.



(r) Diversity plots of wlaB for different noise scale (s) values.

Figure 15: Effect of the CADS parameter noise scale (s) on the Transporters dataset.