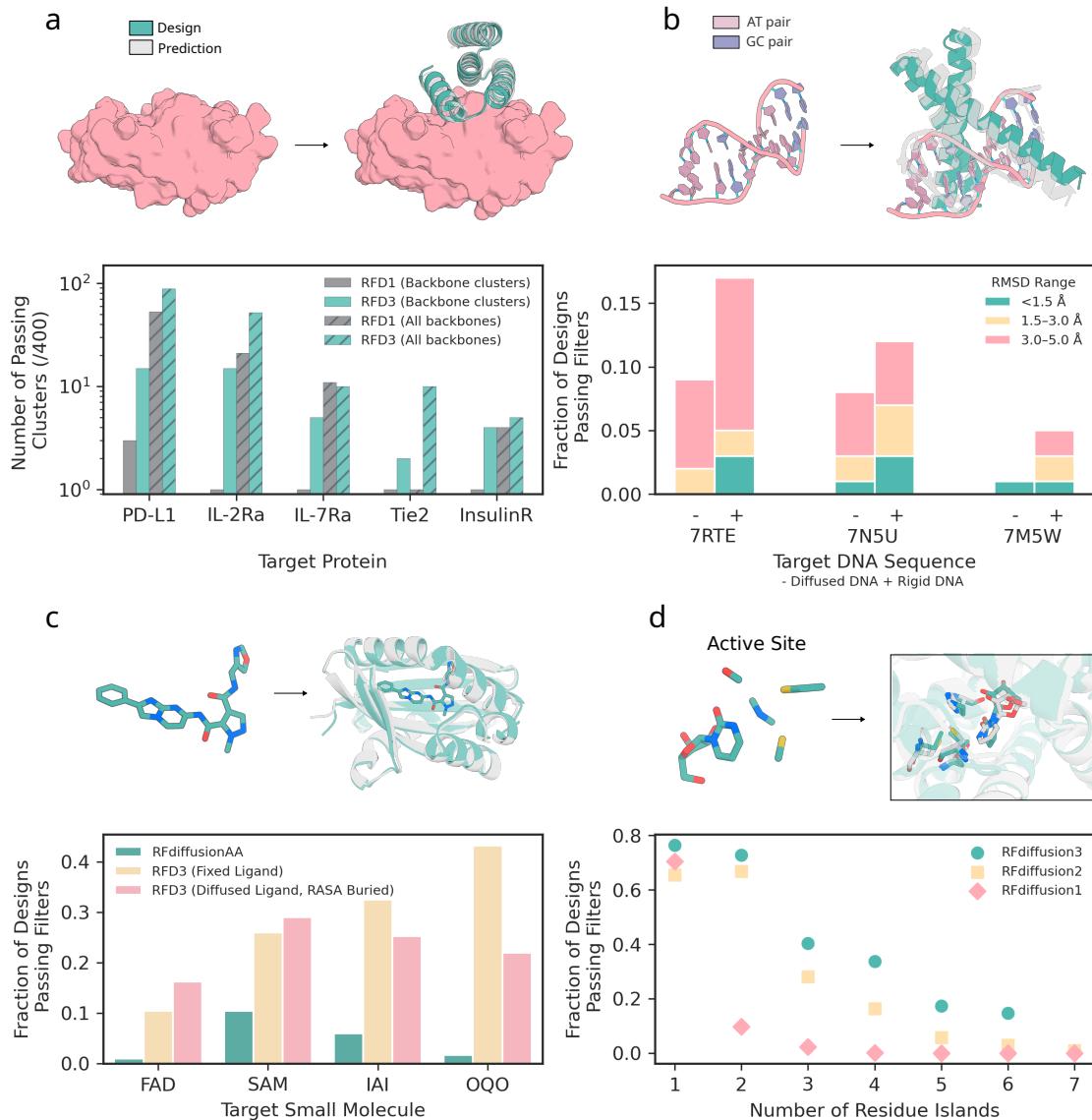


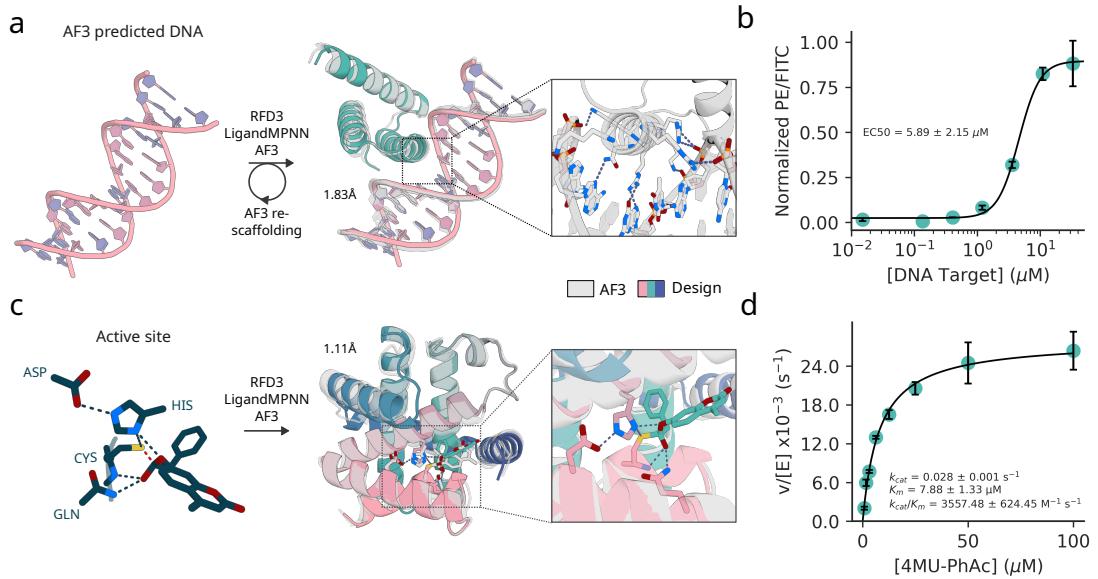
---

**Figure 2 (previous page): Global and atom-level conditioning with RFdiffusion3.** **a**, Diffusion of enzyme active sites. Sidechains are diffused for all residues of the backbone, the optimal position in sequence to scaffold the atomic motif is thereby found during generation, allowing active sites to be scaffolded from simple active site constraints. (top) Zoom in of a trajectory showing the network deciding where in the sequence the side chain residues belong. (bottom) Visualization of predicted denoised structures across different time scales. **b**, Depiction of a diffusion trajectory where both the protein and the DNA are co-generated. In this case, the DNA sequence is given and only the conformation is sampled. **c**, Depiction of a diffusion trajectory where both the protein and the small molecule coordinates are co-generated. In this case, the ligand identity is given and only the conformation is sampled. **d**, Atom-level hydrogen bond conditioning. Both acceptor and donor atoms can be specified as atom level constraints for ligands (top) and nucleic acids (bottom). Applying classifier-free guidance improves adherence to the conditions in both cases. **e**, Atom-level conditioning of solvent accessibility by relative solvent accessible surface area (RASA). (top) example of atomic specification of accessible surface area input specification and output structure (bottom) distributions of the generated structures with different input specifications (over 400 designs from the small molecule benchmark). **f**, The generated centre-of-mass can be guided by the initialized noise cloud. (top) DNA input with two different protein centres of mass. (bottom) generated structures superimposed (with transparency) cluster around the input centre of mass in each case. **g**, Symmetric scaffolds can be specified at inference time by symmetrizing outputs of the diffusion modules. Shown are D2, C3, C5, and C7 all with AF3 C $\alpha$  RMSDs 0.832 Å, 0.450 Å, 0.614 Å, and 0.539 Å respectively.



---

**Figure 3 (previous page): Performance of RFdiffusion3 on *in silico* benchmarks.** All panels: Designs shown in teal, predictions shown in gray (prediction method specified in each figure caption). **a**, Generation of protein binding proteins. RFdiffusion3 was compared to RFdiffusion1 (noise scale 0) across five targets (PD-L1, IL-2Ra, IL-7Ra, Tie2, InsulinR). (top) example generation of protein binding-protein to Tie2 (PDB ID: 2GY5; target structure cropped for visualization) with an AF3 prediction overlayed. (bottom) designs that meet AF3 success (min PAE  $< 1.5$ , pTM binder  $> 0.8$  and RMSD  $< 2.5 \text{ \AA}$  after generating 4 sequences for each backbone with ProteinMPNN) criteria outlined for protein-protein interfaces. Solid bars indicate number of successful clusters (TM score threshold 0.6) and hatched bars represent the total number of successful backbones, both out of 400 total backbones. RFD3 outperforms RFD1 in 4/5 cases without clustering and all cases when clustering, since RFD3 is able to find consistently more diverse solutions to each binding problem. **b**, Generation of DNA-binding proteins. (top) depiction of DNA inputs (PDB ID: 7RTE) and a generated structure with an AF3 prediction overlaid. (bottom) measure of AF3 success (DNA-aligned RMSD) across three DNA targets (PDB ID: 7RTE, 7N5U, 7M5W) across different thresholds (teal  $< 1.5 \text{ \AA}$ , yellow 1.5-3  $\text{ \AA}$ , pink 3-5  $\text{ \AA}$ ). 400 backbones were generated for each DNA target (and 4 sequences fit to each with LigandMPNN) with both fixed DNA conformation from the input PDB (+) or given only the sequences and predicted the DNA shape along with the protein scaffold (-). The minimum RMSD for each backbone was taken as a representative to score each backbone. **c**, Generation of small molecule binding proteins. Four small molecules are benchmarked (two are common in the PDB; FAD, SAM; and two are uncommon; IAI, OQO). (top) Example of an input ligand (IAI) and a generated structure with the AF3 structure overlayed. (bottom) Comparison of RFD3 to previous method RFdiffusionAA on small molecule generation (400 backbones for each condition; backbones for RFdiffusionAA from published work [2]). AF3 success is defined as backbone aligned ligand RMSD  $< 5 \text{ \AA}$ , backbone aligned backbone RMSD  $< 1.5 \text{ \AA}$ , min chain pair PAE  $< 1.5$  and iPTM  $> 0.8$ . Fraction of designs indicates how many of the 400 generated backbones had at least one sequence that passed the criteria described. Results for RFdiffusionAA shown in teal, RFD3 with fixed ligand using a PDB conformation in yellow and RFD3 with diffused ligand and RASA conditioning (described in Figure 2) in pink. **d**, Generation of enzymes. (top) Example scaffolding of a 4-residue active site (M0097; PDB ID: 1CTT). (bottom) Performance across the Atomic Motif Enzyme (AME) benchmark with increasing number of residue islands. Residue islands are defined as the number of contiguous regions in sequence space that are provided as input to scaffold. Passing backbones are defined as those where at least one of 8 LigandMPNN sequences have a Chai-1 predicted motif backbone-aligned motif all-atom RMSD  $< 1.5 \text{ \AA}$ .



**Figure 4: Experimental evaluation of RFdiffusion3.** **a**, Schematic representation of DNA binding protein design with RFdiffusion3. Binders were diffused against a fixed DNA structure of the target sequence CGAGAACATAGTCG and LigandMPNN was used to design sequences. Major groove recognition regions of AF3 passing designs were re-scaffolded to increase diversity. Full details available at supplementary section Section 4.2. Structure shown for validated design DBRFD3, polar interaction in the major groove are highlighted (AF3 prediction). **b**, Normalized binding signal (PE/FITC) from flow cytometry data of yeast surface display based titration assay (without avidity) revealed a low micromolar binding affinity of DBRFD3 (1 out 5 designs tested against the target sequence). The average and standard deviation of the EC<sub>50</sub> value (computed by fitting a four-parameter logistic regression model) is reported based on three replicates. **c**, (left to right) input active site for the cysteine hydrolase design, design structure overlayed with AF3 model of active site geometry. **d**, Michaelis-Menten kinetics of designed Cysteine Hydrolase.

## References

- [1] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023.
- [2] Rohith Krishna, Jue Wang, Woody Ahern, Pascal Sturmels, Preetham Venkatesh, Indrek Kalvet, Gyu Rie Lee, Felix S Morey-Burrows, Ivan Anishchenko, Ian R Humphreys, et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science*, 384(6693):eadl2528, 2024.
- [3] Woody Ahern, Jason Yim, Doug Tischer, Saman Salike, Seth M Woodbury, Donghyo Kim, Indrek Kalvet, Yakov Kipnis, Brian Coventry, Han Raut Altae-Tran, et al. Atom level enzyme active site scaffolding using RFdiffusion2. *bioRxiv*, pages 2025–04, 2025.
- [4] Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, Yehlin Cho, Kourosh H. Ghahary, Laura Vinué, Brahm J. Yachnin, Andrew M. Wollacott, Stephen Buckley, Adrie H. Westphal, Simon Lindhoud, Sandrine Georgeon, Casper A. Goverde, Georgios N. Hatzopoulos, Pierre Gönczy, Yannick D. Muller, Gerald Schwank, Daan C. Swarts, Alex J. Vecchio, Bernard L. Schneider, Sergey Ovchinnikov, and Bruno E. Correia. One-shot design of functional protein binders with BindCraft. *Nature*, 2025. URL <https://doi.org/10.1038/s41586-025-09429-6>.
- [5] Jason Yim, Andrew Campbell, Emile Mathieu, Andrew Y. K. Foong, Michael Gastegger, Jose Jimenez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S. Veeling, Frank Noe, Regina Barzilay, and Tommi Jaakkola. Improved motif-scaffolding with SE(3) flow matching. In *ICLR 2024 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2024. URL <https://openreview.net/forum?id=96JxNcA9r6>.
- [6] Anthony Marchand, Stephen Buckley, Arne Schneuing, Martin Pacesa, Maddalena Elia, Pablo Gainza, Evgenia Elizarova, Rebecca M. Neeser, Pao-Wan Lee, Luc Reymond, Yangyang Miao, Leo Scheller, Sandrine Georgeon, Joseph Schmidt, Philippe Schwaller, Sebastian J. Maerkli, Michael Bronstein, and Bruno E. Correia. Targeting protein–ligand neosurfaces with a generalizable deep learning tool. *Nature*, 639(8054):522–531, 2025. doi: 10.1038/s41586-024-08435-4.
- [7] John B. Ingraham, Max Baranov, Zak Costello, Karl W. Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M. Lord, Christopher Ng-Thow-Hing, Erik R. Van Vlack, Shan Tie, Vincent Xue, Sarah C. Cowles, Alan Leung, João V. Rodrigues, Claudio L. Morales-Perez, Alex M. Ayoub, Robin Green, Katherine Puentes, Frank Oplinger, Nishant V. Panwar, Fritz Obermeyer, Adam R. Root, Andrew L. Beam, Frank J. Poelwijk, and Gevorg Grigoryan. Illuminating Protein Space with a Programmable Generative Model. *Nature*, 623(7989):1070–1078, 2023. URL <https://doi.org/10.1038/s41586-023-06728-8>.
- [8] Vinicius Zambaldi, David La, Alexander E Chu, Harshnira Patani, Amy E Danson, Tristan OC Kwan, Thomas Frerix, Rosalia G Schneider, David Saxton, Ashok Thillaisundaram,

et al. De novo design of high-affinity protein binders with AlphaProteo. *arXiv preprint arXiv:2409.08022*, 2024.

- [9] Jue Wang, Sidney Lisanza, David Juergens, Doug Tischer, Joseph L. Watson, Karla M. Castro, Robert Ragotte, Amijai Saragovi, Lukas F. Milles, Minkyung Baek, Ivan Anishchenko, Wei Yang, Derrick R. Hicks, Marc Expòsit, Thomas Schlichthaerle, Jung-Ho Chun, Justas Dauparas, Nathaniel Bennett, Basile I. M. Wicky, Andrew Muenks, Frank DiMaio, Bruno Correia, Sergey Ovchinnikov, and David Baker. Scaffolding protein functional sites using deep learning. *Science*, 377(6604):387–394, 2022. doi: 10.1126/science.abn2100. URL <https://www.science.org/doi/abs/10.1126/science.abn2100>.
- [10] Chai Discovery Team, Jacques Boitreaud, Jack Dent, Danny Geisz, Matthew McPartlon, Joshua Meier, Zhuoran Qiao, Alex Rogozhnikov, Nathan Rollins, Paul Wollenhaupt, and Kevin Wu. Zero-shot antibody design in a 24-well plate. *bioRxiv*, 2025. doi: 10.1101/2025.07.05.663018. URL <https://www.biorxiv.org/content/early/2025/07/06/2025.07.05.663018>.
- [11] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, pages 1–3, 2024.
- [12] Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, et al. Proteina: Scaling flow-based protein structure generative models. *arXiv preprint arXiv:2503.00710*, 2025.
- [13] Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-Proteina: Atomistic Protein Generation via Partially Latent Flow Matching, 2025. arXiv preprint.
- [14] Alex Bridgland, Jonathan Crabbé, Henry Kenlay, Daniella Pretorius, Sebastian M. Schmon, Agrin Hilmkil, Rebecca Bartke-Croughan, Robin Rombach, Michael Flashman, Tomas Matteson, Simon Mathis, Alexander W. R. Nelson, David Yuan, Annette Obika, and Simon A. A. Kohl. Latent-x: An atom-level frontier model for de novo protein binder design. Technical report, Latent Labs, London and San Francisco, July 2025. URL <https://www.latentlabs.com/wp-content/uploads/2025/07/Latent-X-Technical-Report.pdf>. Technical report. Also available as arXiv:2507.19375.
- [15] Nathaniel Corley, Simon Mathis, Rohith Krishna, Magnus S Bauer, Tuscan R Thompson, Woody Ahern, Maxwell W Kazman, Rafael I Brent, Kieran Didi, Andrew Kubaney, Lilian McHugh, Arnav Nagle, Andrew Favor, Meghana Kshirsagar, Pascal Sturmels, Yanjing Li, Jasper Butcher, Bo Qiang, Lars L Schaaf, Raktim Mitra, Katelyn Campbell, Odin Zhang, Roni Weissman, Ian R Humphreys, Qian Cong, Jonathan Funk, Shreyash Sonthalia, Pietro Liò, David Baker, and Frank DiMaio. Accelerating biomolecular modeling with AtomWorks and RF3. Preprint available at bioRxiv, 2025.
- [16] Wei Qu, Jiawei Guan, Rui Ma, kezhai, Weikun.Wu, and Haobo Wang. P(all-atom) Is Unlocking New Path For Protein Design. In *Forty-second International Conference on Machine Learning*, 2025. URL <https://openreview.net/forum?id=yXRixu0ONY>.

- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.
- [18] Artidoro Pagnoni, Ramakanth Pasunuru, Pedro Rodriguez, John Nguyen, Benjamin Muller, Margaret Li, Chunting Zhou, Lili Yu, Jason E. Weston, Luke Zettlemoyer, Gargi Ghosh, Mike Lewis, Ari Holtzman, and Srinivas Iyer. Byte Latent Transformer: Patches Scale Better Than Tokens. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9238–9258, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.453. URL <https://aclanthology.org/2025.acl-long.453/>.
- [19] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *nature*, 596(7873):583–589, 2021.
- [20] Minkyung Baek, Ivan Anishchenko, Ian R Humphreys, Qian Cong, David Baker, and Frank DiMaio. Efficient and accurate prediction of protein structure using RoseTTAFold2. *BioRxiv*, pages 2023–05, 2023.
- [21] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.
- [22] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.
- [23] wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic acids research*, 47(D1):D520–D528, 2019.
- [24] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Kathryn Tunyasuvunakool, Olaf Ronneberger, Russ Bates, Augustin Žídek, Alex Bridgland, et al. AlphaFold 2. *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction*, 2020.
- [25] Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8946–8970. PMLR, 2022.
- [26] J Dauparas, I Anishchenko, N Bennett, H Bai, R J Ragotte, L F Milles, B I M Wicky, A Courbet, R J de Haas, N Bethel, P J Y Leung, T F Huddy, S Pellock, D Tischer, F Chan,

- B Koepnick, H Nguyen, A Kang, B Sankaran, A K Bera, N P King, and D Baker. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022.
- [27] Justas Dauparas, Gyu Rie Lee, Robert Pecoraro, Linna An, Ivan Anishchenko, Cameron Glasscock, and David Baker. Atomic context-conditioned protein sequence design using Lig-andMPNN. *Nature Methods*, pages 1–7, 2025.
  - [28] Karsten D. Householder, Xinyu Xiang, Kevin M. Jude, Arthur Deng, Matthias Obenaus, Steven C. Wilson, Xiaojing Chen, Nan Wang, and Christopher K. Garcia. De novo design and structure of a peptide-centric TCR mimic binding module. *Science*, 389(6758):375–379, 2025. doi: 10.1126/science.abm1581. URL <https://www.science.org/doi/10.1126/science.abm1581>.
  - [29] Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M Jude, Iva Marković, Rameshwar U Kadam, Koen H G Verschueren, Kenneth Verstraete, Scott Thomas Russell Walsh, Nathaniel Bennett, Ashish Phal, Aerin Yang, Lisa Kozodoy, Michelle DeWitt, Lora Picton, Lauren Miller, Eva-Maria Strauch, Nicholas D DeBouver, Allison Pires, Asim K Bera, Samer Halabiya, Bradley Hammerson, Wei Yang, Steffen Bernard, Lance Stewart, Ian A Wilson, Hannele Ruohola-Baker, Joseph Schlessinger, Sangwon Lee, Savvas N Savvides, K Christopher Garcia, and David Baker. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, May 2022.
  - [30] Nathaniel R Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, Frank DiMaio, Steven De Munck, Savvas N Savvides, and David Baker. Improving de novo protein binder design with deep learning. *Nat. Commun.*, 14(1):2625, May 2023.
  - [31] David M Ichikawa, Osama Abdin, Nader Alerasool, Manjunatha Kogenaru, April L Mueller, Han Wen, David O Giganti, Gregory W Goldberg, Samantha Adams, Jeffrey M Spencer, et al. A universal deep-learning model for zinc finger design enables transcription factor reprogramming. *Nature Biotechnology*, 41(8):1117–1129, 2023.
  - [32] Pratiksha I Thakore and Charles A Gersbach. Design, assembly, and characterization of TALE-based transcriptional activators and repressors. In *TALENs: Methods and Protocols*, pages 71–88. Springer, 2016.
  - [33] Cameron J Glasscock, Robert J Pecoraro, Ryan McHugh, Lindsey A Doyle, Wei Chen, Olivier Boivin, Beau Lonnquist, Emily Na, Yuliya Politanska, Hugh K Haddox, et al. Computational design of sequence-specific DNA-binding proteins. *Nature Structural & Molecular Biology*, pages 1–10, 2025.
  - [34] Remo Rohs, Sean M West, Alona Sosinsky, Peng Liu, Richard S Mann, and Barry Honig. The role of DNA shape in protein–DNA recognition. *Nature*, 461(7268):1248–1253, 2009.
  - [35] Tom J Petty, Soheila Emamzadah, Lorenzo Costantino, Irina Petkova, Elena S Stavridi, Jeffery G Saven, Eric Vauthey, and Thanos D Halazonetis. An induced fit mechanism regulates

- p53 DNA binding kinetics to confer sequence specificity. *The EMBO journal*, 30(11):2167–2176, 2011.
- [36] Remo Rohs, Xiangshu Jin, Sean M West, Rohit Joshi, Barry Honig, and Richard S Mann. Origins of specificity in protein-DNA recognition. *Annual review of biochemistry*, 79(1):233–269, 2010.
- [37] Colin W Garvie and Cynthia Wolberger. Recognition of specific DNA sequences. *Molecular cell*, 8(5):937–946, 2001.
- [38] Raktim Mitra, Jinsen Li, Jared M Sagendorf, Yibei Jiang, Ari S Cohen, Tsu-Pei Chiu, Cameron J Glasscock, and Remo Rohs. Geometric deep learning of protein–DNA binding specificity. *Nature Methods*, 21(9):1674–1683, 2024.
- [39] Nathan M Ennist, Shunzhi Wang, Madison A Kennedy, Mariano Curti, George A Sutherland, Cvetelin Vasilev, Rachel L Redler, Valentin Maffeis, Saeed Shareef, Anthony V Sica, Ash Sueh Hua, Arundhati P Deshmukh, Adam P Moyer, Derrick R Hicks, Avi Z Swartz, Ralph A Cacho, Naia Novy, Asim K Bera, Alex Kang, Banumathi Sankaran, Matthew P Johnson, Amala Phadkule, Mike Reppert, Damian Ekiert, Gira Bhabha, Lance Stewart, Justin R Caram, Barry L Stoddard, Elisabet Romero, C Neil Hunter, and David Baker. De novo design of proteins housing excitonically coupled chlorophyll special pairs. *Nat. Chem. Biol.*, 20(7):906–915, 2024.
- [40] Tim Kaminski, Anders Gunnarsson, and Stefan Geschwindner. Harnessing the versatility of optical biosensors for target-based small-molecule drug discovery. *ACS Sens.*, 2(1):10–15, 2017.
- [41] Patarasuda Chaisupa and R Clay Wright. State-of-the-art in engineering small molecule biosensors and their applications in metabolic engineering. *SLAS Technol.*, 29(2):100113, 2024.
- [42] N. F. Polizzi and W. F. DeGrado. A defined structural unit enables de novo design of small-molecule-binding proteins. *Science*, 369:1227–1233, 2020. doi: 10.1126/science.abb8330.
- [43] Yehlin Cho, Martin Pacesa, Zhidian Zhang, Bruno E Correia, and Sergey Ovchinnikov. Boltzdesign1: Inverting all-atom structure prediction model for generalized biomolecular binder design. *bioRxiv*, pages 2025–04, 2025.
- [44] Keriann M Backus, Bruno E Correia, Kenneth M Lum, Stefano Forli, Benjamin D Horning, Gonzalo E González-Páez, Sandip Chatterjee, Bryan R Lanning, John R Teijaro, Arthur J Olson, Dennis W Wolan, and Benjamin F Cravatt. Proteome-wide covalent ligand discovery in native biological systems. *Nature*, 534(7608):570–574, 2016.
- [45] Lin Jiang, Eric A Althoff, Fernando R Clemente, Lindsey Doyle, Daniela Rothlisberger, Alexandre Zanghellini, Jasmine L Gallaher, Jamie L Betker, Fujie Tanaka, Carlos F Barbas III, et al. De novo computational design of retro-aldol enzymes. *science*, 319(5868):1387–1391, 2008.

- [46] Sarah L Lovelock, Rebecca Crawshaw, Sophie Basler, Colin Levy, David Baker, Donald Hilvert, and Anthony P Green. The road to fully programmable protein catalysis. *Nature*, 606(7912):49–58, 2022.
- [47] Euan J Hossack, Florence J Hardy, and Anthony P Green. Building enzymes through design and evolution. *ACS Catalysis*, 13(19):12436–12444, 2023.
- [48] Anna Lauko, Samuel J. Pellock, Kiera H. Sumida, Ivan Anishchenko, David Juergens, Woody Ahern, Jihun Jeung, Alexander F. Shida, Andrew Hunt, Indrek Kalvet, Christoffer Norn, Ian R. Humphreys, Cooper Jamieson, Rohith Krishna, Yakov Kipnis, Alex Kang, Evans Brackenbrough, Asim K. Bera, Banumathi Sankaran, K. N. Houk, and David Baker. Computational design of serine hydrolases. *Science*, 388(6744):eadu2454, 2025. doi: 10.1126/science.adu2454. URL <https://www.science.org/doi/abs/10.1126/science.adu2454>.
- [49] Alexandre Zanghellini, Lin Jiang, Andrew M Wollacott, Gong Cheng, Jens Meiler, Eric A Althoff, Daniela Röthlisberger, and David Baker. New algorithms and an in silico benchmark for computational enzyme design. *Protein Science*, 15(12):2785–2794, 2006.
- [50] Markus Braun, Adrian Tripp, Morakot Chakatok, Sigrid Kaltenbrunner, Massimo Totaro, David Stoll, Aleksandar Bijelic, Wael Elaily, Shlomo Yakir Hoch, Matteo Aleotti, et al. Computational design of highly active de novo enzymes. *bioRxiv*, pages 2024–08, 2024.
- [51] Chai Discovery, Jacques Boitraud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhnikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *bioRxiv*, pages 2024–10, 2024.
- [52] Harold A Chapman, Richard J Riese, and Guo-Ping Shi. Emerging roles for cysteine proteases in human biology. *Annual review of physiology*, 59(1):63–88, 1997.
- [53] Sven Panke, Martin Held, and Marcel Wubbolt. Trends and innovations in industrial biocatalysis for the production of fine chemicals. *Current opinion in Biotechnology*, 15(4):272–279, 2004.
- [54] Karl Hult and Per Berglund. Engineered enzymes for improved organic synthesis. *Current opinion in biotechnology*, 14(4):395–400, 2003.
- [55] Elena Mossessova and Christopher D. Lima. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Molecular Cell*, 5(5):865–876, 2000. doi: 10.1016/S1097-2765(00)80326-3.
- [56] Florian Richter, Rebecca Blomberg, Sagar D Khare, Gert Kiss, Alexandre P Kuzin, Adam JT Smith, Jasmine Gallaher, Zbigniew Pianowski, Roger C Helgeson, Alexej Grjasnow, et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *Journal of the American Chemical Society*, 134(39):16197–16206, 2012.
- [57] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R. Dustin Schaeffer, Claudia Milán, Hahnbeom Park, Carlos Adams, Craig R Glassman, Alexander DeGiovanni, Jose H