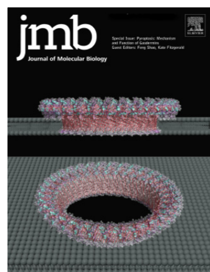
Review Article



Adventures on the Routes of Protein Evolution—In Memoriam Dan Salah Tawfik (1955–2021)[1] ☆

Colin Jackson [1,2]*, Agnes Toth-Petroczy [3,4], Rachel Kolodny [5], Florian Hollfelder [6], Monika Fuxreiter [7], Shina Caroline Lynn Kamerlin [8]* and Nobuhiko Tokuriki [9]*

1 - ***Research School of Chemistry,*** *Australian National University, Canberra, ACT 2601, Australia*
2 - ***Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science,*** *Australian National University, Canberra, ACT 2601, Australia*
3 - ***Max Planck Institute of Molecular Cell Biology and Genetics,*** *01307 Dresden, Germany*
4 - ***Center for Systems Biology Dresden,*** *01307 Dresden, Germany*
5 - ***Department of Computer Science,*** *University of Haifa, Haifa, Israel*
6 - ***Department of Biochemistry,*** *University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK*
7 - ***Department of Biomedical Sciences,*** *University of Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy*
8 - ***Science for Life Laboratory,*** *Department of Chemistry – BMC, Uppsala University, BMC Box 576, S-75123 Uppsala, Sweden*
9 - ***Michael Smith Laboratories,*** *University of British Columbia, Vancouver, BC, V6T 1Z4, Canada*

Correspondence to Colin Jackson, Shina Caroline Lynn Kamerlin and Nobuhiko Tokuriki: Research School of Chemistry, Australian National University, Canberra, ACT, 2601, Australia; Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Australian National University, Canberra, ACT 2601, Australia (C. Jackson); Science for Life Laboratory, Department of Chemistry – BMC, Uppsala University, BMC Box 576, S-75123 Uppsala, Sweden (S.C.L. Kamerlin); Michael Smith Laboratories, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada (N. Tokuriki). *colin.jackson@anu.edu.au (C. Jackson), monika.fuxreiter@unipd.it (M. Fuxreiter), l.kamerlin@gmx.com (S.C.L. Kamerlin), tokuriki@msl.ubc.ca (N. Tokuriki) @Jackson_Lab (C. Jackson), @agnestothp (A. Toth-Petroczy), @rachelkolodny (R. Kolidny), @hollfelderlab (F. Hollfelder), @kamerlinlab (S.C.L. Kamerlin), @tokuriki lab (N. Tokuriki)*
https://doi.org/10.1016/j.jmb.2022.167462
Edited by P. Wright

Abstract

Understanding how proteins evolved not only resolves mysteries of the past, but also helps address challenges of the future, particularly those relating to the design and engineering of new protein functions. Here we review the work of Dan S. Tawfik, one of the pioneers of this area, highlighting his seminal contributions in diverse fields such as protein design, high throughput screening, protein stability, fundamental enzyme-catalyzed reactions and promiscuity, that underpin biology and the origins of life. We discuss the influence of his work on how our models of enzyme and protein function have developed and how the main driving forces of molecular evolution were elucidated. The discovery of the rugged routes of evolution has enabled many practical applications, some which are now widely used.

© 2022 Elsevier Ltd. All rights reserved.

Introduction

"*How do proteins evolve*?" This fundamental question in the life sciences has long fascinated many scientists. Proteins and enzymes are central pieces of biological systems, and the study of the evolutionary principles of proteins requires a broad knowledge in chemistry, biochemistry, evolution and ecology. In the early days, researchers in the area of biochemistry and chemistry largely

focused on understanding the function of extant proteins and enzymes using traditional biochemical and biophysical approaches. However, key technological developments (high throughput screening methodologies, next-generation sequencing, diverse approaches for directed evolution, etc.) in the last decades have generated a *renaissance* of "evolutionary biochemistry", not only by enabling researchers to generate novel proteins and enzymes for diverse applications, but also by allowing researchers to probe models on evolutionary dynamics and theories in the laboratory. In other words, supplementing the study of proteins for what they are with the ability to understand how they came to be through natural evolution and can be manipulated in the future through directed evolution. Moreover, the accumulation of massive amounts of sequence data over recent years has enabled us to explore and infer evolutionary relationships between genes and organisms that have been separated for billions of years. Finally, the development of diverse biochemical and molecular biology approaches (e.g., genome editing, transcriptomics and proteomics) has aided the study of the evolution of proteins in the context of real organisms and environmental selection pressures.

Dan S. Tawfik (1955–2021) was one of the pioneers of evolutionary biochemistry and led many of the key technological and conceptual innovations that have deepened our understanding of the evolution of proteins and enzymes. His own evolution as a scientist, constantly exploring new topics, technologies, and disciplines, mirrored this topic that he was so intrigued by. In this review, we follow his scientific trajectory across rugged evolutionary landscapes, by highlighting a series of his contributions to the seemingly disparate, but connected, fields in which he became involved (Figure 1).

Creating new enzymes – catalytic antibodies to designer enzymes

Although well-known for his work in understanding the molecular evolution of proteins and enzymes, it is a less well-known fact that Dan was a physical organic chemist by training, and it was this foundation in physical organic chemistry that underpinned his molecular-level view of proteins, especially enzymes. Dan was fascinated by enzymes and their ability to efficiently catalyze remarkably diverse chemical reactions. He pursued his PhD under the supervision of Profs. Zelig Eshhar and Michael Sela. For a fascinating and personal recent reflection from Dan on this time in his career and the influence of his training under Eshhar and Sela, we recommend a wonderful article he wrote in the Israeli Journal of Chemistry in 2019.[1] Dan's PhD work focused on the identification and characterization of catalytic antibodies as mimics for enzymes,[2–5] working at an interface between physical organic chemistry and immunology.[6] This tendency to take techniques and expertise from one field and apply them to systems in another apparently unrelated field would repeat throughout his career. While the results were interesting and novel, he was quick to understand the limitations of antibodies as efficient catalysts.[7] Rate accelerations increased linearly with the effective binding energies (K_M/K_i^{TSA}) but rarely beyond,[8] suggesting the lack of reactive groups, catalytically beneficial conformational dynamics, multi-step pathways or other mechanistic complexity. When '*strategic use of haptenic charge*'[9] was proffered as a way out of this dilemma, the discovery of "off-the-shelf proteins" (such as serum albumins) that could compete with the best catalytic antibody described at the time (for the catalysis of the Kemp elimination, a prominent model reaction) suggested that medium effects rather than perfect positioning of the catalytic base may be responsible for catalysis.[10] For a carboxylate base (as in the antibody), an enormous medium effect of 10^8-fold (by non-specific desolvation) has to be considered, and potentially obscures the positioning effect (while the reactivity of an amine base in serum albumins would be paradoxically less dependent on being in a hydrophobic site). Benkovic commented "*the key is in the pocket*".[11] Taking a lead from this perspective, subsequent work on combinatorially modified "synzymes", *i.e.* water-soluble polymers containing amine groups, were shown to act as highly efficient catalysts of Kemp elimination,[12] chemically anticipating the combinatorial approach that was later used to explore enzyme libraries in later work.

His work on catalytic antibodies paved the way for his subsequent participation in a number of landmark *de novo* enzyme design studies, through fruitful collaborations with, among others, David Baker and Ken Houk.[13,14] This work involved taking minimal models of an active site, grafting them onto pre-existing protein scaffolds (using Rosetta atomistic design calculations), and then optimizing the activity further by directed evolution. He was also among the earliest researchers to appreciate the potential of loop grafting and exploiting conformational dynamics for protein engineering,[15] something that is only more recently becoming appreciated by the broader community.[16,17]

Contrary to conventional enzymology at the time, which was predominantly focused on the characterization of extant natural enzymes, his approach to study enzymes involved something of a reverse engineering approach in which enzymes were evolved in the laboratory and the changes investigated in detail, allowing the key traits and mechanisms underlying efficient catalysts to be revealed.[18,19]

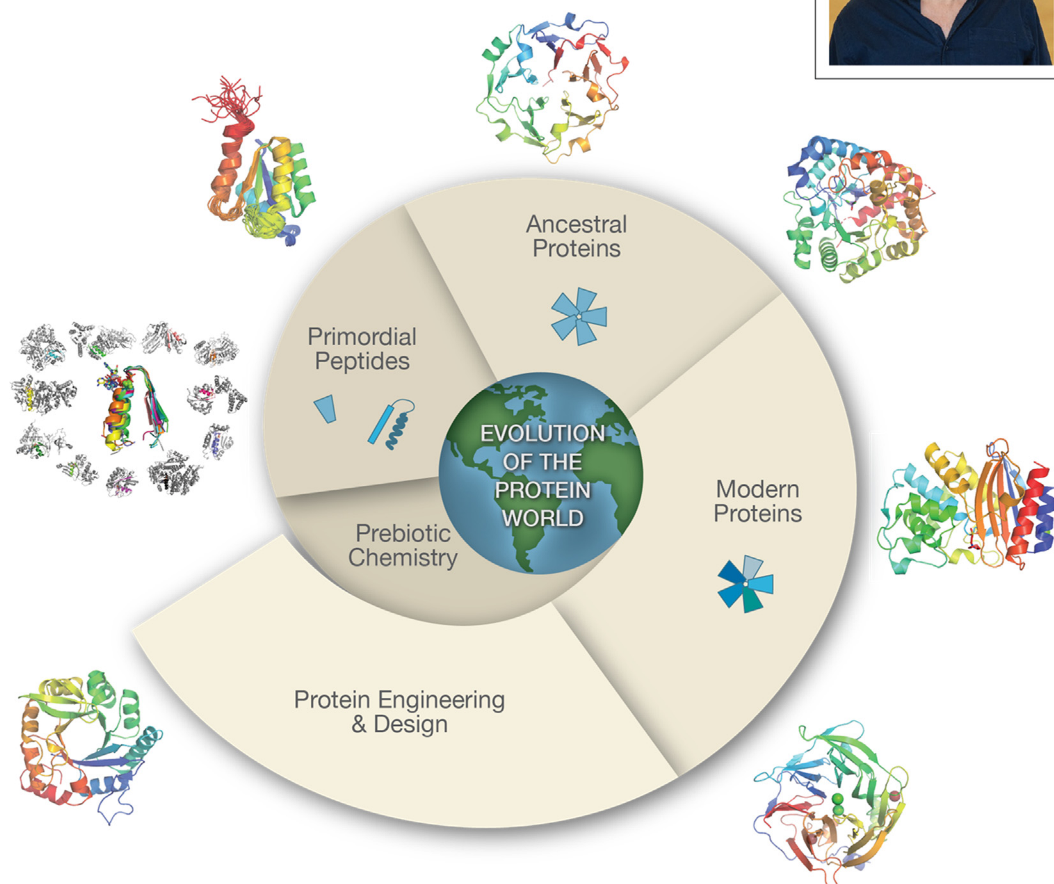


Figure 1. Dan S. Tawfik (1955–2021), a physical organic chemist, who became a protein engineer and later turned into an evolutionist. He has led many of the key technological and conceptual innovations that have deepened our understanding of the evolution of proteins and enzymes. Understanding the fundamental rules of how modern proteins evolve enabled innovations in protein engineering and design. The breadth of his work covered not only the evolution of modern proteins but also primordial proteins and prebiotic chemistry, the chemistry of life. Highlighted are some of what might have been his favorite proteins that he studied and engineered during his career (from left: P-loop design (PDB: 6c2u), tachylectin-2 (PDB: 5c2m), phosphotriesterase (PDB: 4gy1), TEM-1 (PDB: 4ibr), PON (PDB: 4hho), Kemp-eliminase (PDB: 3iio)).

High-throughput screening systems for enzymes – in vitro compartmentalization technology

Through his work on catalytic antibodies, Dan realized the necessity for the technological development of high-throughput directed evolution and protein engineering tools. At the time, several high-throughput technologies had been developed for the selection of protein binders, e.g., phage display and ribosome display, but there were no effective high-throughput methodologies available for screening of enzyme variants (selection/survival was possible but came with its own limitation).

Together with Andrew Griffith, then a fellow in the MRC Laboratory for Molecular Biology, Dan set out to develop an ultrahigh-throughput screening system to allow for the screening of libraries and directly monitor for catalytic turnover. The key idea was to miniaturize test tubes and liquid handling systems into self-assembled reaction vessels, following early work of emulsion droplets technologies.[20–22] The droplet boundary would constitute a link between the genotype and phenotype and constitute an evolutionary unit with μm dimensions and pL volume of which $>10^9$ could fit in a test

tube.[23] The first experiments were entirely cell-free, using single genes that were encapsulated in polydisperse droplets, from which *in vitro* expression systems produced enzyme directly in each compartment.[23] In some ways the emulsion droplets were reminiscent of cells, in terms of size as well as the fact that they contained genetic information and a selectable trait, thereby anticipating minimalist synthetic biology. As applied to directed evolution, this approach allowed selection for DNA self-modification enzymes[24,25] (inspiring Holliger's polymerase evolution in the same format,[26] hydrolases (using one of the first *in vitro* bead display systems)[27] or protein–protein interactions.[28] Dan's group further developed a double emulsion format which made polydisperse droplets amenable to screening for fluorescent reaction products in flow cytometers.[29] Pim Stemmer predicted that "*one day all enzymes will be evolved like this*".[30] This day has not quite come, but the subsequent automation of the droplet generation processes in microfluidic devices has provided an avenue to production (at >kHz rates) of *monodisperse* droplets that are uniform screening compartments for directed evolution[31] or metagenomic enzyme discovery,[32] with assays for more than 20 reaction types available.[33] The impact of *in vitro* compartmentalization reaches beyond enzyme screening: emulsion PCR,[34] digital PCR,[35] transcriptomic analysis (e.g. in the 10x Chromium system[36]) have taken advantage of emulsion droplets as viable experimental formats. Yet, just when the impact of *in vitro* compartmentalization started to resonate, Dan left this fledgling field – quite literally – to its own devices. While screening large numbers of variants may always be better, even the most efficient technology must be combined with a strategy for the exploration of sequence space if the number of clones that can be screened remains lower than the total possible diversity.

Catalytic promiscuity – An evolutionary springboard for new functions

The study of the bovine serum albumin (BSA)-catalyzed reaction and catalytic antibodies led Dan to realize that many proteins and enzymes likely can perform functions that they have never been selected for. The Kemp elimination reaction is not a naturally occurring reaction, thus no protein (including BSA) has ever been under selective pressure to evolve the ability to catalyze Kemp elimination. This observation came as his interest was piqued in the obscured side of enzyme specificity - "catalytic promiscuity" or versatility, i.e., ability of enzymes that catalyze secondary and unselected chemical reactions.[37] Enzymes have evolved under selective pressure to catalyze chemical reactions with specific substrate(s) *via* certain chemical reactions in the active site. However, while characteristically specific, enzyme active sites cannot (in general) absolutely exclude all other potential substrates, especially those that have not previously existed in natural environments (synthetic chemicals) for which there has been no selective pressure against. Consequently, catalytic promiscuity is likely an inherent property of most, if not all, enzymes in Nature. After Dan moved back to the Weizmann Institute of Science as a principal investigator, his group demonstrated that catalytic promiscuity is indeed a ubiquitous feature of many enzymes by exploring diverse enzymes[38] including mammalian paraoxonases (PONs),[39–41] carbonic anhydrase II (CAII),[42] and microbial lactonases and phosphotriesterases.[43,44] Moreover, Dan built upon previous inspiration from Jensen,[37] and molecular conceptualisation by Herschlag,[45] that catalytic promiscuity has played a significant role in evolution of diverse chemistry and substrate specificity in natural enzymes.[46,47] His groups contribution to the field in "playing out" how catalytic promiscuity can affect evolution, through directed evolution, was a major contribution to the area and allowed the hypothesis to be rigorously tested. This demonstrated an existing, albeit low-level, promiscuous activity can provide a springboard for evolution, which bridges different chemical reactions and enzymes. Numerous subsequent studies further demonstrated that indeed the diversity of natural enzyme specificity has been underpinned by the expansion of pre-existing promiscuous activities of ancestral enzymes.[46–51]

The link between enzyme promiscuity, protein dynamics and evolvability

Dan also explored the molecular basis of functional promiscuity, or multi-functionality, and the evolutionary mechanisms of protein functions. His pioneering work with Leo James demonstrated that multi-specific antibodies sample alternative conformations, facilitating the recognition of unrelated antigens.[52–54] The kinetic data of these binding studies were compatible with a model that included an induced fit isomerisation step between two alternative conformers upon binding to a hapten. However, structural diversity may also permit a wider substrate spectrum than required (cross-reactivity), leading to unwanted side-effects, like autoimmunity. These results are consistent with the model[55] that sequences of proteins may encode multiple structures and multiple functions including functional promiscuity, in contrast to the deterministic structure–function paradigm.[56,57] His subsequent studies provided experimental evidence for the link between conformational dynamics, multi-functionality and catalytic promiscuity in many enzyme systems.[58–63]

Along with exploring catalytic promiscuity in many enzymes, Dan's group performed a number of

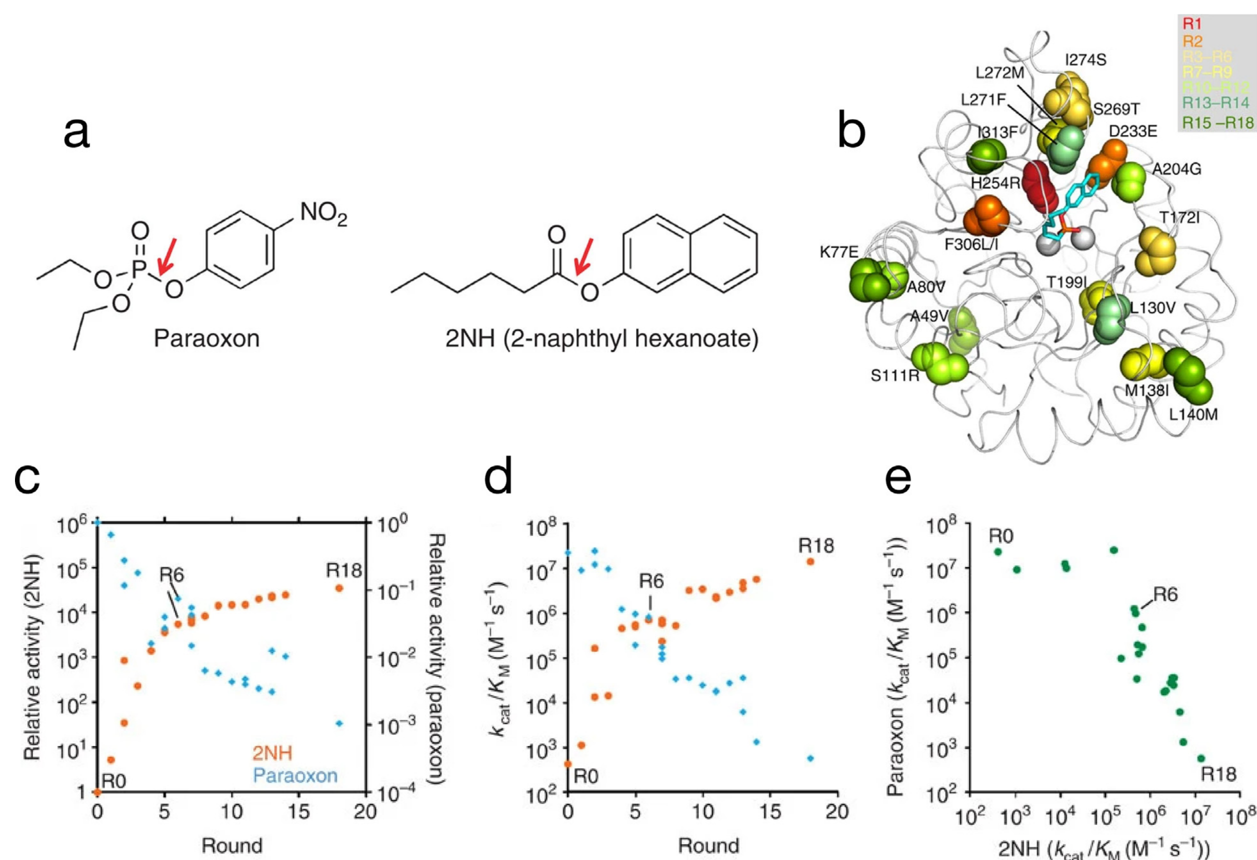


Figure 2. An example of a "long" *in vitro* evolutionary trajectory from Tokuriki et al.[61] In this example, a phosphotriesterase was evolved to hydrolase aryl esters (a). This exemplifies many of the concepts of molecular evolution that were developed by Dan and colleagues. For example, (b) we observe the mutations begin within the active site, then become progressively more remote throughout the trajectory, as the conformational dynamics of the protein are changed. (c) Catalytic promiscuity is seen as the two activities overlap and very gradually change in terms of the relative activities across the trajectory. (d) Stability and activity interact epistatically, as function-changing but destabilizing mutations were incorporated with the use of chaperones (e.g. R2-5), while other generations (e.g. R8) involved incorporation of stabilizing mutations had no benefit to the catalytic activity but increased whole cell activity (c) by increasing soluble expression. (e) Finally, we observe what are now characteristic evolutionary trajectories marked by a weak-negative trade-off, where the new activity is gained without commensurate loss of the original activity in the early stages, followed by diminishing returns in the improvement over the final generations of optimization. This figure was adapted from Ref. 61, and is published under a CC-BY 4.0 license.

directed evolution experiments that enhanced existing promiscuous activities[28,38,41,42,61,62,64–66] (for a comprehensive review, see ref.[46]). These provided evidence for a scenario characterised by weak trade-offs[44] (Figure 2(e)): promiscuous functions can easily improve with a handful of mutations, yet those mutations tend to exhibit negligible impact on the original function. He equated tolerant molecular recognition with protein 'evolvability'[39]: if a native activity can tolerate mutations leading to a second activity,[67] the emerging generalist enzymes will support both original and new catalytic reactions.[61] This scenario illustrated how gene duplication during the divergence of enzymes can rapidly lead to a head start for the new function, providing molecular recognition contingency for a rapid response to a new selection pressure for the promiscuous function.

Eventually, Dan integrated concepts of functional promiscuity, conformational dynamics, and protein evolvability, and developed an integrated 'New View' of protein evolution.[57] In this model, conformational diversity promotes functional promiscuity. Protein conformational space can be conceptualized as an ensemble of multiple distinct conformational states. Some dominant conformational states may exclusively perform the native and physiological function while other minor conformers may fit to other, promiscuous substrates.[56] Evolution to improve promiscuous functions may occur by shifting the population balance between distinct conformational and functional substates. Subsequent

directed evolution studies supported the model that conformational variability and functional diversity are linked in proteins, expanding the biological repertoire of the same sequence.[68] This proposal was consistent with earlier results suggesting that robustness of cellular processes may confer evolutionary benefits with precise structural requirements.

As previously discussed, Dan's group demonstrated that alternative conformations can be further facilitated by increased dynamics or local disorder, which can promote reorganization of active sites and otherwise destabilizing function-related mutations.[59] At the same time, excessive disorder can also lead to non-productive conformations, via perturbing the interaction network required for a given enzymatic activity.[60] Moreover, functional variability may not only come from changes in protein structures.[69] Indeed, there are a wide range of other components, such as errors in gene expression that result in phenotypic mutations,[70] including frameshifts,[71,72] and changes in expression levels or patterns of post-translational modifications, which contribute to alteration of phenotypes.[70,73] In a broader cellular context, the protein–protein interaction network, local concentrations of metabolites or signaling molecules, and the higher-order organization of cellular components can also contribute to phenotypic noise.[74] In an intriguing commentary, Dan proposed that such messiness could be a source, or even the origin, of many biological innovations.[69]

Stability constraints in protein evolution

Through a series of directed evolution studies, Dan also noticed that the laboratory evolution of many proteins stagnated and that it quickly became impossible improve their functions further after several rounds of directed evolution, while for some proteins it was not possible to improve their function at all through directed evolution.[75] By carefully tracing evolutionary trajectories generated in the lab, Dan realized that protein stability was a major bottleneck for functional evolution. As proteins are often marginally stable, any mutations that lower protein stability will be purged out even if the mutation is functionally beneficial.[76] To probe this, his group conducted a systematic survey for the effect of mutations on protein stability and function by characterizing mutations obtained from a number of directed evolution studies,[77–79] demonstrating that function-altering mutations are destabilizing on average, and that the accumulation of multiple mutations leads to unstable protein variants that cannot tolerate further mutations. Such function-stability trade-offs,[76] and the often epistatic relationship between them,[80,81] is now recognised as one of the major evolutionary dynamics in most proteins and enzymes.

The question then naturally became: How have proteins circumvented this evolutionary stability bottleneck and continuously expanded their functional repertoire in Nature? This question has both fundamental[82] and applied implications: while Dan sought to understand how proteins overcome stability constraints in protein evolution, the solutions and strategies that his group discovered could be implemented in protein engineering. One approach is to simply incorporate stabilizing mutations to compensate for the destabilizing effect of function-altering mutations.[80] His group has explored and developed various methods to identify stabilizing mutations, such as incorporation of ancestral and consensus mutations,[83,84] and performing neutral drift to increase diversity prior to selection.[67] Another way is to utilize chaperones, such as GroEL/ES, to buffer the destabilizing effect of function-altering mutations.[65] Overexpression of chaperones can increase the capacity of the cell to assist the folding of diverse proteins, including partially destabilized enzyme mutants that carry potential function-improving mutations, and thus chaperones can increase evolvability of proteins. With the combination of those stability modulating strategies, it is now possible to perform long-term enzyme engineering to generate new enzymes with native-like efficiency in the laboratory[61,85] (Figure 2).

Protein innovability

The in-depth study of several key dynamics in protein evolution led Dan to further develop an integrated view of protein evolvability; specifically, that protein evolvability includes two conflicting elements - robustness or neutrality i.e., mutations having no or little effect on fitness, and innovability i.e., mutations readily inducing new functions, or innovations. He was intrigued by how these two conflicting demands are bridged[79,86,87]; how has Nature circumvented stability-innovability trade-offs and generated diverse functions? What are the key molecular elements in the protein architecture to support innovability?

Dan's group performed a large-scale bioinformatics analysis and observed that the functional diversity within each protein fold, and therefore its innovability, differ substantially depending on the scaffold.[86] Certain folds, such as TIM barrels and Rossman folds, accommodate hundreds of diverse functions with no sequence similarity, whereas other folds, such as dihydrofolate reductase evolved only one enzymatic function despite existing for billions of years since the last universal common ancestor (LUCA). His group further explored the molecular basis for enzyme innovability.[86,87] They have proposed that fold polarity is a key factor, which is high if the active site is composed of flexible, loosely packed loops that are well-separated from the highly ordered scaffold.

Folds with high polarity tend to exhibit high innovability. For example, the TIM barrel fold shows high scaffold–active-site polarity (∼20% of the active site comprises scaffold residues) and >2-fold higher rates of sequence divergence at active-site positions. Similarly, antibodies that Dan studied early in his career exhibit high polarity and represent an optimal fold for engineering,[2,3] readily allowing, for example, loop grafting[15]; an endeavour that is largely unattainable in other proteins. On the contrary, folds in which the active site is embedded within the scaffold, e.g. DHFR, tend to exhibit narrower spectrum of functional diversity.[86] In non-polar folds, the active-site residues cannot diverge unless the scaffold does, and vice versa. In general, correlation between residues, or epistasis, dramatically slows down the rate of evolution because a change in one residue is dependent on a change in other residues.[88–91]

Taken together, modularity contributes to both robustness and innovability, for example in the assembly of multi-domain proteins[92] or in case of folds, where a high fraction of active-site residues are weakly connected to the scaffold.[87]

Understanding the chemistry of life

Dan was an avid reader of scientific papers and a true scientist, and he loved to test hypotheses, both his own and of others. His interest in natural evolution, which deepened into a fascination with primordial chemistry and the origin of life was, to a degree, sparked by the controversial study that suggested the GFAJ-1 microbe from Mono Lake in California could incorporate arsenic in place of phosphorus in DNA,[93] and later studies that refuted this finding.[94,95] He rapidly began studying the physicochemical properties of arsenic and phosphorus and how they can be enzymatically substituted for each other.[96] Again, his background in physical organic chemistry came to the fore, as well as his interest in the PhD work of a postdoc in the laboratory at the time, Mikael Elias, on phosphate binding proteins.[97] With Mikael, Dan quickly used this system to test the hypothesis of whether arsenate is incorporated into the biology of GFAJ-1 through transport across the membrane by solute binding proteins. Using a combination of biophysical approaches, including protein crystallography, they showed that binding of arsenate was strongly disfavored by all the putative solute binding transporter systems, elegantly demonstrating that, despite their superficial similarity, a network of dipole-anion and repulsive interactions results in arsenate distorting a very short hydrogen bond.[98] This means that even though the thermochemical radii of phosphate and arsenate differ by only 4%, the binding proteins have evolved to discriminate against incorporation of arsenate by many orders of magnitude even in the presence of excess arsenate. Thus, this work stands as another example where Dan used his vast experience and knowledge of chemistry address an outstanding problem in biology.

Following from this, he turned his interest to other examples of natural evolution. For instance, the question of where the vast amount of dimethyl sulfide in the oceans comes from. This work was a long collaboration with the group of Assaf Vardi, an expert in marine algae. Together, their labs first identified and characterized an enzyme called DddD, demonstrating that dimethylsulfoniopropionate is its native substrate.[99] This fundamental work was then quickly followed by a series of influential papers that have dramatically changed our understanding of this compound and its role in biology and even our atmosphere. From the discovery of the alma1 DMS lyase gene family in phytoplankton,[100] to developing specific alma lyase inhibitors to allow identification of DMS sources[101], and finally elucidating its role in signalling and how it can mediate prey-predator dynamics in the aquatic ecosystem in work that was published posthumously only this year.[102]

His fascination with the chemistry of life and evolution is also illustrated in his work on chalcone isomerase[103]; a plant enzyme that catalyzes the stereospecific isomerization of chalcones in the biosynthesis of anthocyanins. This was a nice coincidence of interests from the chemistry of life to a longstanding question in enzymology: how powerful catalysts can arise from non-catalytic scaffolds. Following the discovery that chalcone isomerase could have plausibly evolved from a fatty acid binding protein by the group of Joseph Noel,[104] Dan's group collaborated with Joseph, to show that seemingly dramatic changes in function, such as the acquisition of catalytic activity, can be achieved through gradual stepwise reshaping of the active site of a binding protein and repositioning of a key catalytic arginine side chain in the active site.[58] Furthermore, curiously, this arginine appears to play an important dual role as both a Brønsted and a Lewis acid in the chalcone-isomerase catalyzed reaction (Figure 3).[105]

Most recently, Dan derived a timeline of biosphere oxygenation from occurrence patterns of oxygen-utilizing enzymes.[106] Typically, geological data was used for this, placing the Great Oxidation Event, when oxygen became a permanent feature of the atmosphere, at ∼2.4 billion years ago.[107,108] In a very different approach, he mapped the emergence of O_2 enzyme families to significant events in the species tree: the LUCA-node (Last Universal Common Ancestor – the separation of archaea and bacteria), the LUOA (Last Universal Oxygen Ancestor)-node (separating terrestrial and marine bacteria), or later.[106] Because these are very ancient events, they carefully focused on families that were not involved in early horizontal gene transfer (HGT) and compared the expansion rates to baseline rates derived from other enzyme fami-

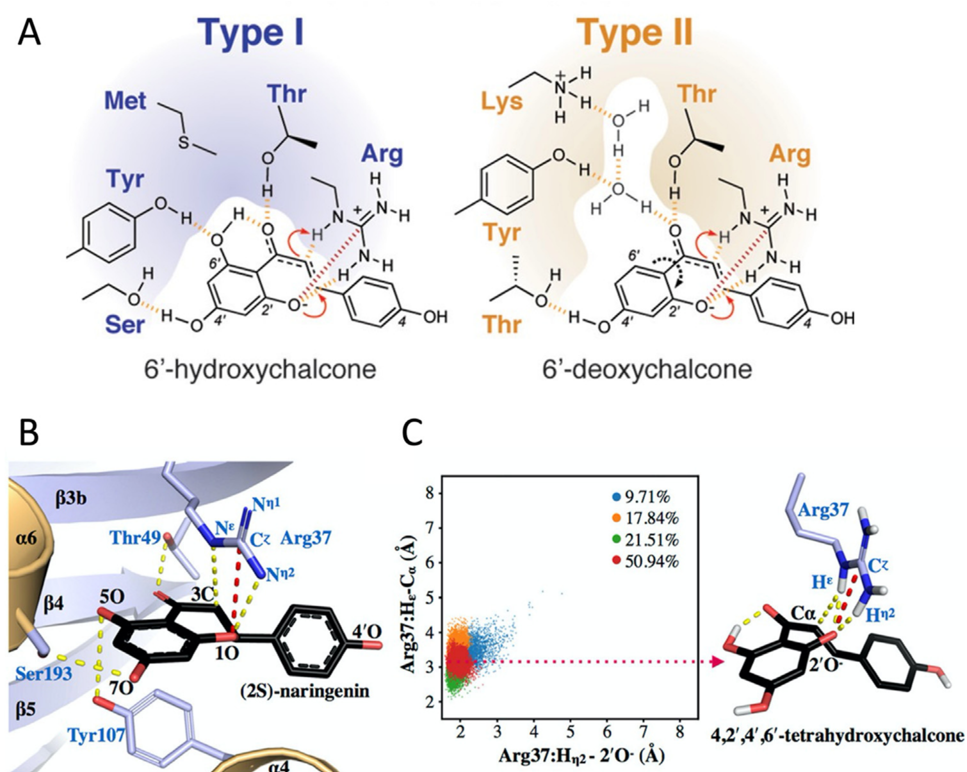


Figure 3. (A) Comparison of the active site architectures of Type I and Type II chalcone isomerases, in complex with 6′-hydroxychalcone and 6′-deoxychalcone, respectively, highlighting the position of the catalytic arginine side chain. (B) Examination of enzyme-product interactions in the *Mt*CH1 active site (in complex with product (2*S*)-naringenin) highlight how the Arg37 side chain activates the substrate through both hydrogen bonding (dotted yellow lines) and a Lewis acid (dotted red line) interaction, thus facilitating ring opening. (C) Exploration of the conformational space of the Arg37 side chain in molecular dynamics interactions (distance definitions highlighted in the figure in the inset), showing the sampling of stable and catalytically favorable enzyme substrate interactions between the Arg37 side chain and the substrate 4,2′,4′,6′-tetrahydroxychalone. This figure was adapted from Ref. 105, and is reproduced with permission. Copyright 2019, American Chemical Society.

lies. Observing that most oxygen-utilizing enzymes emerged at the LUOA, an evolutionary event dated to ∼3.1 billion years ago, they concluded that oxygen was widely available far earlier than proposed.

The origin of life: Primordial peptides and ancestral proteins

Following his work understanding the chemistry of our contemporary world, Dan became increasingly interested in deciphering ancient protein history, including that before domains emerged. He appreciated the elegant and seminal work of Dayhoff,[109] and in a paper celebrating her achievements he also outlined work he set out to do.[110] It was postulated that ancient proteins were likely shorter peptides, possibly from only a subset of amino acids. One can therefore search for "fossils" of such ancient polypeptides in internally symmetric proteins or in omnipresent proteins that originated in the LUCA. Short polypeptides, not embedded in globular proteins, may also function via self-assembly. His group used theoretical and experimental approaches to study these questions. Their theoretical analyses carefully reasoned about the ancient world based on patterns in current day proteins, while ruling out alternative scenarios that could have led to these patterns.[106,111–114] Experimentally, his group used protein engineering to study ancient and feasible evolutionary trajectories.[74,115,116]

A repeated thread in his theoretical analyses was differentiating between divergent and convergent evolution. He considered both explanations: of similar present-day versions that diverged from a pre-LUCA common ancestor, or proteins that evolved independently under shared functional demands and chemical-physical constraints and converged to similar molecular traits. In cases of ancient, shared ancestry, where the sequence similarity is diminished, he studied alternative meaningful molecular traits. For example, he argued that Rossmann enzymes diverged from a common β2-Asp/Glu motif in a pre-LUCA ancestor.[114] His group further studied the prevalence of phospho-ligand binding within defined families and in different evolutionary lineages[111]; when

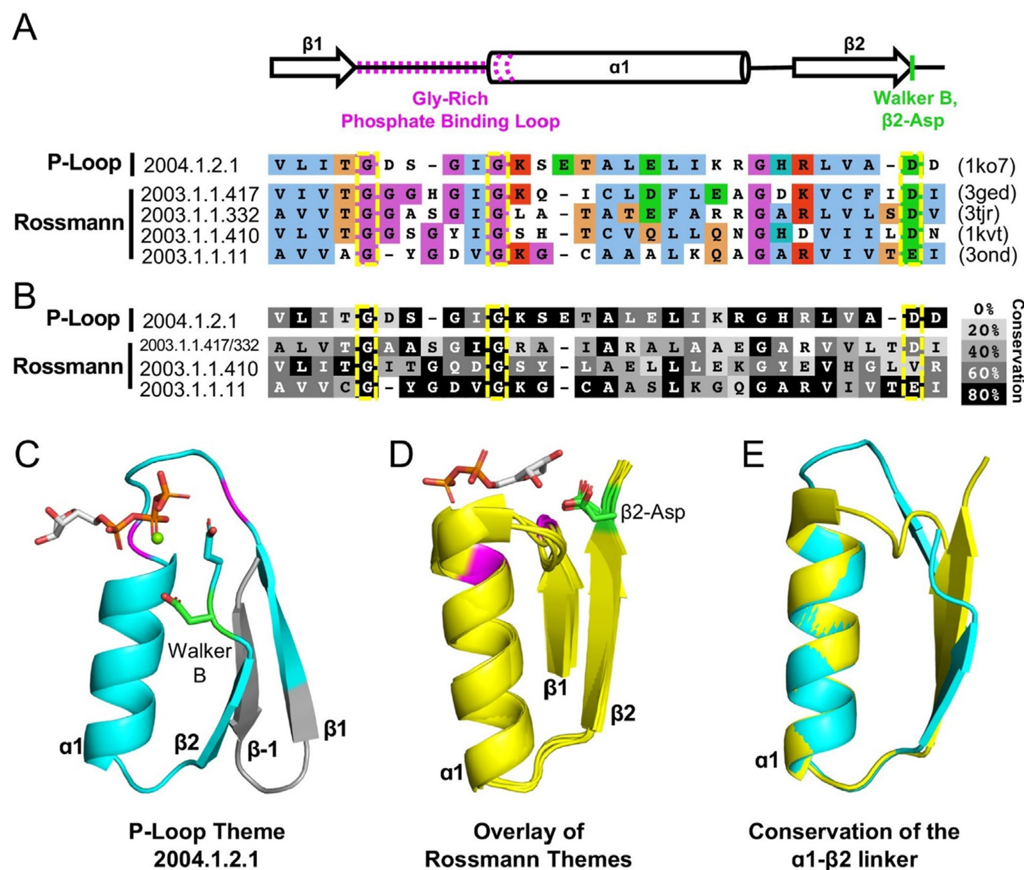


Figure 4. A bridging theme shared between the P-loop (ECOD 2004.1.2.1) and Rossmann fold proteins (ECOD 2003.1.1.{417.332,410,11} is shown: (**A–B**) The sequence alignment of the shared themes found in a segment of a single P-loop NTPase, Hpr kinase (top line, ECOD domain e1ko7A1), and multiple Rossmann fold proteins in four different F-groups, along with the consensus sequence of each F-group is shaded according to the degree of conservation. (**C–D**) The structure of the shared theme encompasses the β1-PBL-α1-β2-Asp element in both the P-loop protein (panel **C**; Hpr kinase, ECOD domain e1ko7A1) and the theme-related Rossmann fold proteins (panel **D**; ECOD domains e3gedA1, e1kvtA1, e3ondA1, and e3tjrA1; the ligand is bound by domain e3tjrA1). (**E**) An overlay of the β1-PBL-α1-β2-Asp element of the Hpr Kinase (cyan; ECOD domain e1ko7A1) and one of the theme-related Rossmann dehydrogenases (yellow; ECOD domain e3tjrA1). This figure was adapted from Ref. 111, and is published under a CC-BY 4.0 license.

widespread within families, he concluded it was a founding function of that lineage, and vice versa for later-added niche function to only some families. Indeed, in the most ancient lineages, including P-loops and the Rossmann fold, Rossmann structures with crossover, TIM-barrels, HUP domains, the flavodoxin fold, and HAD domains, phospho-ligand binding appears to be a founding function. Analyzing the binding mode within these ancient lineages revealed that the mode of N-helix binding, where a β-strand is followed by a phosphate-binding loop and then a flanking α-helix, predates other modes. He argued that this binding mode supports divergent evolution. However, because the evidence was not conclusive, he put forward another, complementary explanation for this shared molecular trait: that the preference for α-helical binding sites in the ancient phosphate binders reflects the constraints acting on the earliest short and prebiotic-sequence proteins.

In search of pre-LUCA ancestral segments, Dan's group studied proteins, which share a significant segment despite belonging to distinct evolutionary lineages, also termed bridge proteins.[113] Unlike previous studies,[117] structural similarity between shared segments was not required, and indeed, in approximately half of the cases found, the structures differ. Dan argued that this structural variation is consistent with sequence homology being the outcome of common ancestry, rather than convergence due to structural constraints. It may also be that the capacity of a peptide to adapt its structure to different environments is evolutionarily advantageous.[56] The set of bridging pairs that were studied included tangible links between some of the most ancient enzyme classes: flavodoxin-like, TIM-barrel, P-loop NTPase and Rossmann-like.

Dan considered the P-loop and Rossmann motifs to be prime suspects in the search for ancestral segments,[110] which prompted his studies of their

evolution[114,116] and of the emergence of their phospho-ligand binding sites.[112] Thus, he was excited to find a link among the bridging proteins connecting P-loops and the Rossmann motif (Figure 4).[111] The two folds have an overall similar architectures and in both, the phosphate binding motif uses the tip of the first helix (α1), even though this is not a biophysical necessity, as evidenced by a few Rossmann families that use another helix (α4).[111] It may appear that the two folds differ significantly due to their different strand order leading to binding in the Rossmann that uses the β2 strand (and an aspartic acid therein), while in P-loops it is typically served by the β3 or β4 strand. Dan argued however, that this difference is reconciled by the bridging P-loop domain (PDB ID:1KO7) which binds like several Rossmann domains to the tip of β2. While carefully cautioning the reader against ruling out convergent evolution, the paper offers new support to the hypothesis that these folds diverged from a common ancestor.[111]

Using protein engineering, Dan's group studied how self-assembly enables short ancient polypeptides to function. He predicted an ancestral P-loop segment by phylogenetic analysis of the Walker-A motif (β-strand followed by a phosphate-binding loop and then a flanking α-helix).[116] Embedding this segment in a short, de novo designed (55 residues) protein scaffold created soluble, stable proteins. For these proteins, oligomerization improves binding to polynucleotides, RNA, and single-stranded DNA. In a series of similar studies, his group explored how structurally diverse folds emerge *de novo*, with a particular emphasis on proteins with high internal structural symmetry, using beta-propellers as a model system. He established that these folds likely evolved from short, functional gene segments that, at later stages, duplicated, fused, and rearranged, to yield the folds we see today.[118] The reconstruction of such putative evolutionary pathways after many millions of years of evolutionary drift is difficult. However, his group demonstrated that a 100 residue long segment could assemble into active homo-pentamers.[119] Later, his group identified even shorter sequence modules taken from tachylectin-2 that comprise three strands of one blade plus one strand of the next blade, thus enabling the closure of the propeller's ring *via* strand-strand velcro-like interactions.[120] Finally, his group phylogenetically reconstructed a 47-residue ancestral motif that forms five-bladed lectin propellers via oligomeric assembly.[115] In addition to expanding our fundamental understanding of protein evolution, this work also provides an important baseline for modular design of new enzymes, an area which has witnessed substantial advances in recent years.

In one of his last experimental studies, Dan constructed a feasible trajectory from a simple polypeptide of abiotic amino acids to a modern protein that binds nucleic acids.[74] A challenge in this study was to engineer a polypeptide that binds nucleic acids, but only with spontaneously synthesized amino acids (as these are assumed to be ancient ones). This subset of amino acids does not include the basic Lys, Arg, and His, which are critical for binding nucleic acids. To solve this, they synthesized a polypeptide that uses ornithine, a basic abiotic amino acid that exists in today's cells, but not in today's proteins. Starting from this initial polypeptide, which had only the assumed-ancient function of forming coacervates with RNA, they demonstrated a plausible evolutionary trajectory that changes it to an avid and selective dsDNA binder.

Conclusions and perspectives

Dan's passion for studying proteins and enzymes was formulated in his early career. Then, he, a physical organic chemist, became a protein engineer and later turned into an evolutionist as he has gradually evolved his research subjects and approaches. Throughout his career, Dan used his broad knowledge to combine different scientific areas, ceaseless curiosity and love of science to continuously explore new technologies, new concepts and even new fields. One principle he extolled was to not let previous work or data constrain your thinking or ideas – to think forward and conceptualise new ideas first, rather than incrementally following the latest trend. He was also passionate about making each paper worthwhile, especially reviews and perspectives, into which he generously poured new ideas and thoughts about various fields. He has bridged traditionally distinct disciplines and developed several key technologies and concepts that have been widely used in the field. His legacy will remain with us and keep inspiring us to challenge ourselves to explore new science.

Received 6 January 2022;
Accepted 17 January 2022;
Available online 29 January 2022

Keywords:
protein evolution;
enzymatic catalysis;
high-throughput screening;
primordial enzymes;
catalytic promiscuity

☆ Dan S. Tawfik was a member of the JMB Editorial Board between 2012 and 2021.

References

1. Tawfik, D.S., (2019). A personal reflection on the chemistry-biology interface. *Isr. J. Chem.* **59**, 23–28.

2. Green, B.S., Tawfik, D.S., (1989). Catalytic monoclonal antibodies: tailor-made, enzyme-like catalysts for chemical reactions. *Trends Biotechnol.* **7**, 304–310.
3. Tawfik, D.S., Green, B.S., Chap, R., Sela, M., Eshhar, Z., (1993). catELISA: a facile general route to catalytic antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 373–377.
4. Tawfik, D.S., Zemel, R.R., Arad-Yellin, R., Green, B.S., Eshhar, Z., (1990). Simple method for selecting catalytic monoclonal antibodies that exhibit turnover and specificity. *Biochemistry* **29**, 9916–9921.
5. Tawfik, D.S., Eshhar, Z., Green, B.S., (1993). From Phosphonates to Catalytic Antibodies. A Novel Route to Phosphonoester Transition State Analogs and Haptens. *Phosphorus Sulfur Silicon Relat. Elem.* **76**, 123–126.
6. Jencks, W.P., (1969). Catalysis in Chemistry and Enzymology. Dover, Mineola, NY.
7. Tawfik, D.S., Eshhar, Z., Green, B.S., (1994). Catalytic antibodies: a critical assessment. *Mol. Biotechnol.* **1**, 87–103.
8. Stewart, J.D., Benkovic, S.J., (1995). Transition-state stabilization as a measure of the efficiency of antibody catalysis. *Nature* **375**, 388–391.
9. Thorn, S.N., Daniels, R.G., Auditor, M.T., Hilvert, D., (1995). Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* **373**, 228–230.
10. Hollfelder, F., Kirby, A.J., Tawfik, D.S., (1996). Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature* **383**, 60–62.
11. Benkovic, S.J., (1996). Antibody catalysis. The key is in the pocket. *Nature* **383**, 23–24.
12. Hollfelder, F., Kirby, A.J., Tawfik, D.S., (1997). Efficient catalysis of proton transfer by synzymes. *J. Am.*. https://doi.org/10.1021/ja971970y.
13. Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., et al., (2008). Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195.
14. Khersonsky, O., Kiss, G., Röthlisberger, D., Dym, O., Albeck, S., Houk, K.N., Baker, D., Tawfik, D.S., (2012). Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10358–10363.
15. Tawfik, D.S., (2006). Loop Grafting and the Origins of Enzyme Species. *Science* **311**, 475–476.
16. Nestl, B.M., Hauer, B., (2014). Engineering of Flexible Loops in Enzymes. *ACS Catal.* **4**, 3201–3211.
17. Crean, R.M., Gardner, J.M., Kamerlin, S.C.L., (2020). Harnessing Conformational Plasticity to Generate Designer Enzymes. *J. Am. Chem. Soc.* **142**, 11324–11342.
18. Peisajovich, S.G., Tawfik, D.S., (2007). Protein engineers turned evolutionists. *Nature Methods* **4**, 991–994.
19. Trudeau, D.L., Tawfik, D.S., (2019). Protein engineers turned evolutionists-the quest for the optimal starting point. *Curr. Opin. Biotechnol.* **60**, 46–52.
20. Lederberg, J., (1954). A simple method for isolating individual microbes. *J. Bacteriol.* **68**, 258–259.
21. Nossal, G.J., Lederberg, J., (1958). Antibody production by single cells. *Nature* **181**, 1419–1420.
22. Rotman, B., (1961). Measurement of activity of single molecules of β-d-galactosidase. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1981–1991.
23. Miller, O.J., Bernath, K., Agresti, J.J., Amitai, G., Kelly, B. T., Mastrobattista, E., Taly, V., Magdassi, S., et al., (2006). Directed evolution by in vitro compartmentalization. *Nature Methods* **3**, 561–570.
24. Tawfik, D.S., Griffiths, A.D., (1998). Man-made cell-like compartments for molecular evolution. *Nature Biotechnol.* **16**, 652–656.
25. Cohen, H.M., Tawfik, D.S., Griffiths, A.D., (2004). Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. *Protein Eng. Des. Sel.* **17**, 3–11.
26. Ghadessy, F.J., Ong, J.L., Holliger, P., (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4552–4557.
27. Griffiths, A.D., Tawfik, D.S., (2003). Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J.* **22**, 24–35.
28. Levin, K.B., Dym, O., Albeck, S., Magdassi, S., Keeble, A. H., Kleanthous, C., Tawfik, D.S., (2009). Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nature Struct. Mol. Biol.* **16**, 1049–1055.
29. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., Tawfik, D.S., (2005). High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* **12**, 1281–1289.
30. Minshull, J., Stemmer, W.P., (1999). Protein evolution by molecular breeding. *Curr. Opin. Chem. Biol.* **3**, 284–290.
31. Kintses, B., Hein, C., Mohamed, M.F., Fischlechner, M., Courtois, F., Lainé, C., Hollfelder, F., (2012). Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. *Chem. Biol.* **19**, 1001–1009.
32. Colin, P.-Y., Kintses, B., Gielen, F., Miton, C.M., Fischer, G., Mohamed, M.F., Hyvönen, M., Morgavi, D.P., et al., (2015). Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nature Commun.* **6**, 1–12.
33. Neun, S., Zurek, P.J., Kaminski, T.S., Hollfelder, F., (2020). Ultrahigh throughput screening for enzyme function in droplets. *Methods Enzymol.* **643**, 317–343.
34. Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S., Griffiths, A.D., (2006). Amplification of complex gene libraries by emulsion PCR. *Nature Methods* **3**, 545–550.
35. Vogelstein, B., Kinzler, K.W., (1999). Digital PCR. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9236–9241.
36. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., et al., (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495.
37. Jensen, R.A., (1976). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425.
38. Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., Tawfik, D.S., (2005). The "evolvability" of promiscuous protein functions. *Nature Genet.* **37**, 73–76.
39. Ben-David, M., Elias, M., Filippi, J.-J., Duñach, E., Silman, I., Sussman, J.L., Tawfik, D.S., (2012). Catalytic versatility and backups in enzyme active sites: the case of serum paraoxonase 1. *J. Mol. Biol.* **418**, 181–196.

40. Khersonsky, O., Tawfik, D.S., (2005). Structure-reactivity studies of serum paraoxonase PON1 suggest that its native activity is lactonase. *Biochemistry* **44**, 6371–6382.
41. Amitai, G., Gupta, R.D., Tawfik, D.S., (2007). Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* **1**, 67–78.
42. Gould, S.M., Tawfik, D.S., (2005). Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* **44**, 5444–5452.
43. Roodveldt, C., Tawfik, D.S., (2005). Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry* **44**, 12728–12736.
44. Afriat, L., Roodveldt, C., Manco, G., Tawfik, D.S., (2006). The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry* **45**, 13677–13686.
45. O'Brien, P.J., Herschlag, D., (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105.
46. Khersonsky, O., Tawfik, D.S., (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471–505.
47. Khersonsky, O., Roodveldt, C., Tawfik, D.S., (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508.
48. Copley, S.D., (2021). Setting the stage for evolution of a new enzyme. *Curr. Opin. Struct. Biol.* **69**, 41–49.
49. Chevrette, M.G., Gutiérrez-García, K., Selem-Mojica, N., Aguilar-Martínez, C., Yañez-Olvera, A., Ramos-Aboites, H.E., Hoskisson, P.A., Barona-Gómez, F., (2020). Evolutionary dynamics of natural product biosynthesis in bacteria. *Nature Prod. Rep.* **37**, 566–599.
50. Baier, F., Copp, J.N., Tokuriki, N., (2016). Evolution of Enzyme Superfamilies: Comprehensive Exploration of Sequence-Function Relationships. *Biochemistry* **55**, 6375–6388.
51. Copley, S.D., (2017). Shining a light on enzyme promiscuity. *Curr. Opin. Struct. Biol.* **47**, 167–175.
52. James, L.C., Roversi, P., Tawfik, D.S., (2003). Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367.
53. James, L.C., Tawfik, D.S., (2005). Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12730–12735.
54. James, L.C., Tawfik, D.S., (2003). The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Sci.* **12**, 2183–2193.
55. Frauenfelder, H., Sligar, S.G., Wolynes, P.G., (1991). The energy landscapes and motions of proteins. *Science* **254**, 1598–1603.
56. James, L.C., Tawfik, D.S., (2003). Conformational diversity and protein evolution–a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **28**, 361–368.
57. Tokuriki, N., Tawfik, D.S., (2009). Protein dynamism and evolvability. *Science* **324**, 203–207.
58. Kaltenbach, M., Burke, J.R., Dindo, M., Pabis, A., Munsberg, F.S., Rabin, A., Kamerlin, S.C.L., Noel, J.P., et al., (2018). Evolution of chalcone isomerase from a noncatalytic ancestor. *Nature Chem. Biol.* **14**, 548–555.
59. Dellus-Gur, E., Elias, M., Caselli, E., Prati, F., Salverda, M.L.M., de Visser, J.A.G.M., Fraser, J.S., Tawfik, D.S., (2015). Negative Epistasis and Evolvability in TEM-1 β-Lactamase–The Thin Line between an Enzyme's Conformational Freedom and Disorder. *J. Mol. Biol.* **427**, 2396–2409.
60. Ben-David, M., Sussman, J.L., Maxwell, C.I., Szeler, K., Kamerlin, S.C.L., Tawfik, D.S., (2015). Catalytic stimulation by restrained active-site floppiness–the case of high density lipoprotein-bound serum paraoxonase-1. *J. Mol. Biol.* **427**, 1359–1374.
61. Tokuriki, N., Jackson, C.J., Afriat-Jurnou, L., Wyganowski, K.T., Tang, R., Tawfik, D.S., (2012). Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Commun.* **3**, 1257.
62. Afriat-Jurnou, L., Jackson, C.J., Tawfik, D.S., (2012). Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling. *Biochemistry* **51**, 6047–6055.
63. Jackson, C.J., Foo, J.-L., Tokuriki, N., Afriat, L., Carr, P. D., Kim, H.-K., Schenk, G., Tawfik, D.S., et al., (2009). Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21631–21636.
64. Khersonsky, O., Rosenblat, M., Toker, L., Yacobson, S., Hugenmatter, A., Silman, I., Sussman, J.L., Aviram, M., et al., (2009). Directed evolution of serum paraoxonase PON3 by family shuffling and ancestor/consensus mutagenesis, and its biochemical characterization. *Biochemistry* **48**, 6644–6654.
65. Tokuriki, N., Tawfik, D.S., (2009). Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673.
66. Rockah-Shmuel, L., Tawfik, D.S., (2012). Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. *Nucleic Acids Res.* **40**, 11627–11637.
67. Bershtein, S., Goldin, K., Tawfik, D.S., (2008). Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029–1044.
68. Campbell, E., Kaltenbach, M., Correy, G.J., Carr, P.D., Porebski, B.T., Livingstone, E.K., Afriat-Jurnou, L., Buckle, A.M., et al., (2016). The role of protein dynamics in the evolution of new enzyme function. *Nature Chem. Biol.* **12**, 944–950.
69. Tawfik, D.S., (2010). Messy biology and the origins of evolutionary innovations. *Nature Chem. Biol.* **6**, 692–696.
70. Goldsmith, M., Tawfik, D.S., (2009). Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6197–6202.
71. Yanagida, H., Gispan, A., Kadouri, N., Rozen, S., Sharon, M., Barkai, N., Tawfik, D.S., (2015). The Evolutionary Potential of Phenotypic Mutations. *PLoS Genet.* **11**, e1005445
72. Rockah-Shmuel, L., Tóth-Petróczy, Á., Sela, A., Wurtzel, O., Sorek, R., Tawfik, D.S., (2013). Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. *PLoS Genet.* **9**, e1003882
73. Rotem, E., Loinger, A., Ronin, I., Levin-Reisman, I., Gabay, C., Shoresh, N., Biham, O., Balaban, N.Q., (2010). Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12541–12546.

74. Longo, L.M., Despotović, D., Weil-Ktorza, O., Walker, M. J., Jabłońska, J., Fridmann-Sirkis, Y., Varani, G., Metanis, N., et al., (2020). Primordial emergence of a nucleic acid-binding protein via phase separation and statistical ornithine-to-arginine conversion. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15731–15739.
75. Roodveldt, C., Aharoni, A., Tawfik, D.S., (2005). Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* **15**, 50–56.
76. Tokuriki, N., Tawfik, D.S., (2009). Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604.
77. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., Tawfik, D.S., (2007). The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* **369**, 1318–1332.
78. Tokuriki, N., Stricher, F., Serrano, L., Tawfik, D.S., (2008). How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002
79. Tokuriki, N., Oldfield, C.J., Uversky, V.N., Berezovsky, I. N., Tawfik, D.S., (2009). Do viral proteins possess unique biophysical features? *Trends Biochem. Sci.* **34**, 53–59.
80. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., Tawfik, D.S., (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929–932.
81. Soskine, M., Tawfik, D.S., (2010). Mutational effects and the evolution of new protein functions. *Nature Rev. Genet.* **11**, 572–582.
82. Rockah-Shmuel, L., Tóth-Petróczy, Á., Tawfik, D.S., (2015). Systematic Mapping of Protein Mutational Space by Prolonged Drift Reveals the Deleterious Effects of Seemingly Neutral Mutations. *PLoS Comput. Biol.* **11**, e1004421
83. Trudeau, D.L., Kaltenbach, M., Tawfik, D.S., (2016). On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. *Mol. Biol. Evol.* **33**, 2633–2641.
84. Alcolombri, U., Elias, M., Tawfik, D.S., (2011). Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. *J. Mol. Biol.* **411**, 837–853.
85. Kaltenbach, M., Jackson, C.J., Campbell, E.C., Hollfelder, F., Tokuriki, N., (2015). Reverse evolution leads to genotypic incompatibility despite functional and active site convergence. *Elife* **4** https://doi.org/10.7554/eLife.06492.
86. Dellus-Gur, E., Toth-Petroczy, A., Elias, M., Tawfik, D.S., (2013). What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J. Mol. Biol.* **425**, 2609–2621.
87. Tóth-Petróczy, A., Tawfik, D.S., (2014). The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* **26**, 131–138.
88. Tóth-Petróczy, Á., Tawfik, D.S., (2013). Protein insertions and deletions enabled by neutral roaming in sequence space. *Mol. Biol. Evol.* **30**, 761–771.
89. Tóth-Petróczy, Á., Tawfik, D.S., (2011). Slow protein evolutionary rates are dictated by surface – core association. *Proc. Natl. Acad. Sci. U. S. A.* **108** https://doi.org/10.1073/pnas.1015994108.
90. Breen, M.S., Kemena, C., Vlasov, P.K., Notredame, C., Kondrashov, F.A., (2012). Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538.
91. Wellner, A., Raitses Gurevich, M., Tawfik, D.S., (2013). Mechanisms of protein sequence divergence and incompatibility. *PLoS Genet.* **9**, e1003665
92. Marsh, J.A., Teichmann, S.A., (2010). How do proteins gain new domains? *Genome Biol.* **11**, 126.
93. Wolfe-Simon, F., Switzer Blum, J., Kulp, T.R., Gordon, G. W., Hoeft, S.E., Pett-Ridge, J., Stolz, J.F., Webb, S.M., et al., (2011). A bacterium that can grow by using arsenic instead of phosphorus. *Science* **332**, 1163–1166.
94. Reaves, M.L., Sinha, S., Rabinowitz, J.D., Kruglyak, L., Redfield, R.J., (2012). Absence of detectable arsenate in DNA from arsenate-grown GFAJ-1 cells. *Science* **337**, 470–473.
95. Erb, T.J., Kiefer, P., Hattendorf, B., Günther, D., Vorholt, J.A., (2012). GFAJ-1 is an arsenate-resistant, phosphate-dependent organism. *Science* **337**, 467–470.
96. Tawfik, D.S., Viola, R.E., (2011). Arsenate replacing phosphate: alternative life chemistries and ion promiscuity. *Biochemistry* **50**, 1128–1134.
97. Liebschner, D., Elias, M., Moniot, S., Fournier, B., Scott, K., Jelsch, C., Guillot, B., Lecomte, C., et al., (2009). Elucidation of the phosphate binding mode of DING proteins revealed by subangstrom X-ray crystallography. *J. Am. Chem. Soc.* **131**, 7879–7886.
98. Elias, M., Wellner, A., Goldin-Azulay, K., Chabriere, E., Vorholt, J.A., Erb, T.J., Tawfik, D.S., (2012). The molecular basis of phosphate discrimination in arsenate-rich environments. *Nature* **491**, 134–137.
99. Alcolombri, U., Laurino, P., Lara-Astiaso, P., Vardi, A., Tawfik, D.S., (2014). DddD is a CoA-transferase/lyase producing dimethyl sulfide in the marine environment. *Biochemistry* **53**, 5473–5475.
100. Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D.S., Vardi, A., (2015). MARINE SULFUR CYCLE. Identification of the algal dimethyl sulfide-releasing enzyme: A missing link in the marine sulfur cycle. *Science* **348**, 1466–1469.
101. Alcolombri, U., Lei, L., Meltzer, D., Vardi, A., Tawfik, D.S., (2017). Assigning the Algal Source of Dimethylsulfide Using a Selective Lyase Inhibitor. *ACS Chem. Biol.* **12**, 41–46.
102. Shemi, A., Alcolombri, U., Schatz, D., Farstey, V., Vincent, F., Rotkopf, R., Ben-Dor, S., Frada, M.J., et al., (2021). Dimethyl sulfide mediates microbial predator-prey interactions between zooplankton and algae in the ocean. *Nature Microbiol.* **6**, 1357–1366.
103. Jez, J.M., Bowman, M.E., Dixon, R.A., Noel, J.P., (2000). Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nature Struct. Biol.* **7**, 786–791.
104. Ngaki, M.N., Louie, G.V., Philippe, R.N., Manning, G., Pojer, F., Bowman, M.E., Li, L., Larsen, E., et al., (2012). Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* **485**, 530–533.
105. Burke, J.R., La Clair, J.J., Philippe, R.N., Pabis, A., Corbella, M., Jez, J.M., Cortina, G.A., Kaltenbach, M., et al., (2019). Bifunctional Substrate Activation via an Arginine Residue Drives Catalysis in Chalcone Isomerases. *ACS Catal.* **9**, 8388–8396.
106. Jabłońska, J., Tawfik, D.S., (2021). The evolution of oxygen-utilizing enzymes suggests early biosphere oxygenation. *Nature Ecol. Evol.* **5**, 442–448.

107. Holland, H.D., (2006). The oxygenation of the atmosphere and oceans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 903–915.
108. Holland, H.D., (2002). Volcanic gases, black smokers, and the great oxidation event. *Geochim. Cosmochim. Acta* **66**, 3811–3826.
109. Barnabas, J., Schwartz, R.M., Dayhoff, M.O., (1982). Evolution of major metabolic innovations in the precambrian. *Orig. Life* **12**, 81–91.
110. Romero Romero, M.L., Rabin, A., Tawfik, D.S., (2016). Functional Proteins from Short Peptides: Dayhoff's Hypothesis Turns 50. *Angew. Chem. Int. Ed Engl.* **55**, 15966–15971.
111. Longo, L.M., Jabłońska, J., Vyas, P., Kanade, M., Kolodny, R., Ben-Tal, N., Tawfik, D.S., (2020). On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment. *Elife* **9** https://doi.org/10.7554/eLife.64415.
112. Longo, L.M., Petrović, D., Kamerlin, S.C.L., Tawfik, D.S., (2020). Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5310–5318.
113. Kolodny, R., Nepomnyachiy, S., Tawfik, D.S., Ben-Tal, N., (2021). Bridging Themes: Short Protein Segments Found in Different Architectures. *Mol. Biol. Evol.* **38**, 2191–2208.
114. Laurino, P., Tóth-Petróczy, Á., Meana-Pañeda, R., Lin, W., Truhlar, D.G., Tawfik, D.S., (2016). An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. *PLoS Biol.* **14**, e1002396
115. Smock, R.G., Yadid, I., Dym, O., Clarke, J., Tawfik, D.S., (2016). De Novo Evolutionary Emergence of a Symmetrical Protein Is Shaped by Folding Constraints. *Cell* **164**, 476–486.
116. M.L. Romero Romero, F. Yang, Y.-R. Lin, A. Toth-Petroczy, I.N. Berezovsky, A. Goncearenco, W. Yang, A. Wellner, F. Kumar-Deshmukh, M. Sharon, D. Baker, G. Varani, D.S. Tawfik, Simple yet functional phosphate-loop proteins, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E11943–E11950.
117. Alva, V., Söding, J., Lupas, A.N., (2015). A vocabulary of ancient peptides at the origin of folded proteins. *Elife* **4**, e09410
118. Yadid, I., Tawfik, D.S., (2007). Reconstruction of functional beta-propeller lectins via homo-oligomeric assembly of shorter fragments. *J. Mol. Biol.* **365**, 10–17.
119. Yadid, I., Kirshenbaum, N., Sharon, M., Dym, O., Tawfik, D.S., (2010). Metamorphic proteins mediate evolutionary transitions of structure. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7287–7292.
120. Yadid, I., Tawfik, D.S., (2011). Functional β-propeller lectins by tandem duplications of repetitive units. *Protein Eng. Des. Sel.* **24**, 185–195.