

A “Fuzzy”-Logic Language for Encoding Multiple Physical Traits in Biomolecules

Shira Warszawski, Ravit Netzer, Dan S. Tawfik and Sarel J. Fleishman

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence to Sarel J. Fleishman: sarel@weizmann.ac.il

<http://dx.doi.org/10.1016/j.jmb.2014.10.002>

Edited by S. A. Teichmann

Abstract

To carry out their activities, biological macromolecules balance different physical traits, such as stability, interaction affinity, and selectivity. How such often opposing traits are encoded in a macromolecular system is critical to our understanding of evolutionary processes and ability to design new molecules with desired functions. We present a framework for constraining design simulations to balance different physical characteristics. Each trait is represented by the equilibrium fractional occupancy of the desired state relative to its alternatives, ranging from none to full occupancy, and the different traits are combined using Boolean operators to effect a “fuzzy”-logic language for encoding any combination of traits. In another paper, we presented a new combinatorial backbone design algorithm *AbDesign* where the fuzzy-logic framework was used to optimize protein backbones and sequences for both stability and binding affinity in antibody-design simulation. We now extend this framework and find that fuzzy-logic design simulations reproduce sequence and structure design principles seen in nature to underlie exquisite specificity on the one hand and multispecificity on the other hand. The fuzzy-logic language is broadly applicable and could help define the space of tolerated and beneficial mutations in natural biomolecular systems and design artificial molecules that encode complex characteristics.

© 2014 MRC Laboratory of Molecular Biology. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

Introduction

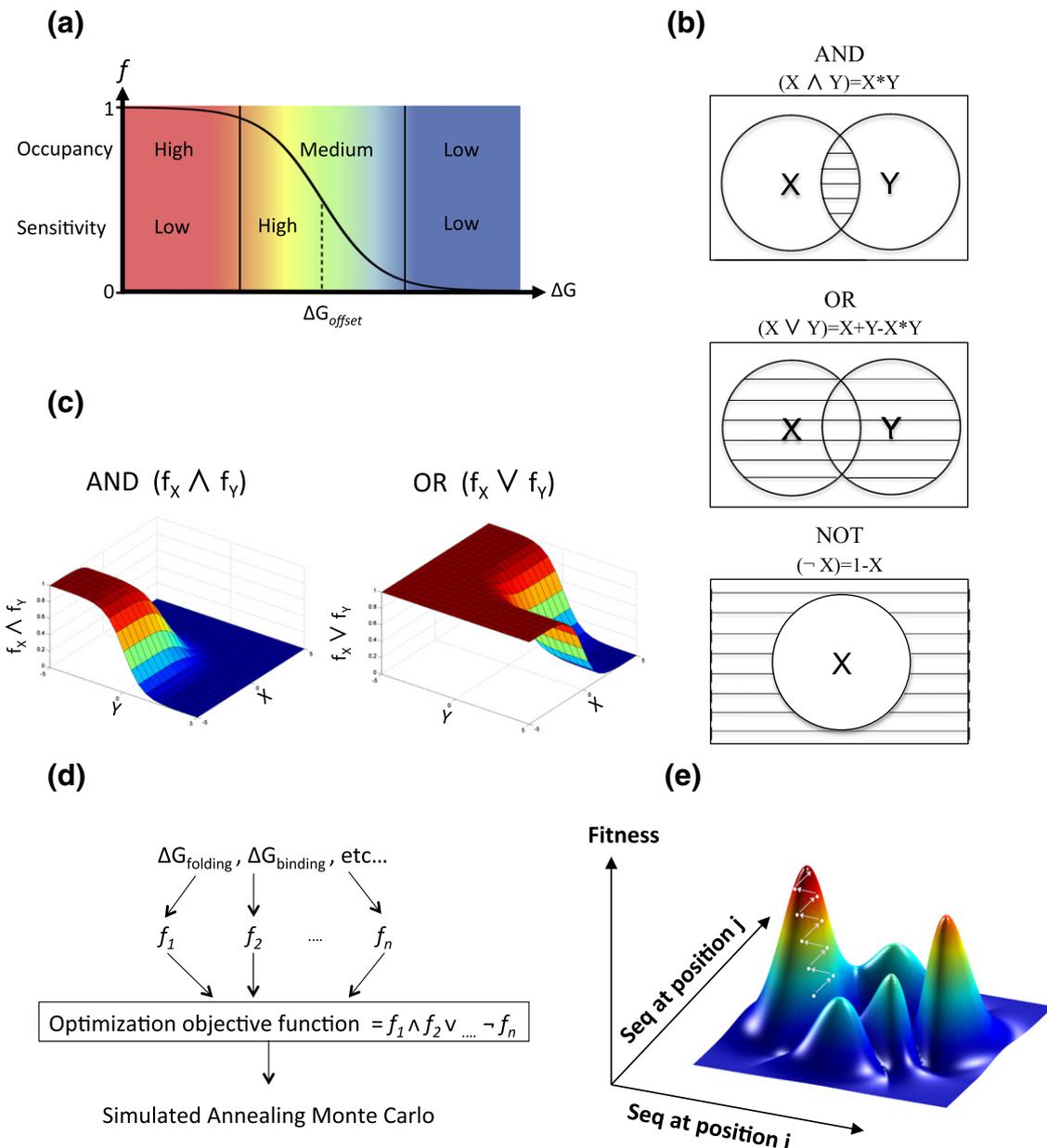
The physical properties of biological macromolecules are subject to evolutionary selection to improve organism fitness. In all cases, more than one biomolecular property is under selection, and complex tradeoffs between competing traits often constrain the evolutionary optimization process. As a very general example, all enzymes need to fold into stable three-dimensional structures in order to function. Therefore, all enzyme sequences, regardless of their functional class, encode favorable free-energy changes for folding [1–4], as well as for their catalytic function, which itself presents a complex tradeoff among substrate, transition state, and product binding affinities [5]. Stability, selectivity, and higher catalytic efficiency are often opposing objectives in the selection of biomolecules, such as proteins or RNAs. Thus, mutations that improve stability often come at the price of reduced binding affinity [6] and catalytic efficiency [7–12] and vice versa [13]. Biomolecular specificity,

where an enzyme or receptor recognizes a set of substrates or ligands while excluding noncognate molecules of similar structure, is often encoded by features, including sequence and backbone changes, that destabilize undesirable contacts even at the expense of lower affinity or catalytic efficiency for the desired partners [14,15]. Conversely, multispecific enzymes and receptors exhibit similar affinities for a range of ligands rather than the highest possible affinity for one [16]. These and many other examples demonstrate the universal importance of encoding multiple constraints and tradeoffs for understanding and engineering biomolecular systems [15].

By controlling all inputs into the process, computational protein design could provide a way to rigorously formulate and test our understanding of the physical constraints that shape biological molecules and particularly of tradeoffs between these constraints [17]. To date, however, most design strategies explicitly maximized one desired physical property through a strategy broadly known as positive design [18], where the

sequence and structure are optimized to lower the energy of the target state. Positive design strategies have led, among other applications, to the design and experimental validation of thermostabilized protein variants [19,20], a highly thermostable novel protein fold [21], idealized natural folds [22], enzymes [23,24], and protein binders [25–28]. However, the strategy of maximizing molecular traits stands to reason when only one molecular trait, such as stability, is optimized [19–22] or when molecular traits do not tradeoff, that is, come at the expense of one another. When tradeoffs apply, maximizing any subset of properties can have the undesired consequence of disfavoring sequences that could simultaneously optimize all desired traits, at least to sufficiency [29].

Computational design methods that explicitly encode multiple constraints have been suggested to design and analyze specificity and multispecificity. These methods consider several different states for the designed molecule in a strategy broadly known as multistate design. A particularly influential concept has been to maximize the energy gap between the target (positive) state and undesired (negative) states to generate pairs of molecules that interact more favorably with one another than with homologous natural partners [30–32]. Conversely, multispecificity has been analyzed by computationally optimizing the sum of energies computed for complexes formed between a hub protein and its ligands, rather than maximizing the binding energy between any particular



pair [33]. Recent progress in *de novo* binder design relied on optimizing the designed binders' rigidity at the price of losing favorable contacts with the target molecule, thus encoding some elements of the biomolecular tradeoff [34–38]. While these strategies share in common the concept of encoding multiple physical traits and tradeoffs between them, they use fundamentally different analytical frameworks and heuristics, each tailored to the particular design objective under consideration.

Here, we develop a general framework for optimizing biomolecular systems operating under conditions where tradeoffs emerge. Our framework is based on equilibrium-thermodynamics principles

and is supported by recent observations on the natural and laboratory evolution of biomolecules. In another paper, we used this new framework to encode both stability and binding in the design of antibody binders (Lapidot et al., unpublished results). Here, we develop the theoretical grounds for this framework, generalize it to more complex relationships between physical traits (including tradeoffs between constraints), and subject it to two classic tests of multiconstrained design: the design of protein-binding specificity and multispecificity. We find that designed complexes qualitatively recapitulate the energy and structure features observed in their natural counterparts.

Results

Quantifying physical traits using fractional occupancies

We consider cases where evolutionary selection operates on molecular properties that are under chemical equilibrium such as folding, binding, or conformational changes. Most generally, we can describe these properties as an equilibrium between two states, A and B, where one state, say B, exhibits the desirable molecular function and thereby promotes “fitness”. For example, A and B could represent, respectively, the unfolded and folded states of a protein or an RNA molecule, the unbound and bound states of a receptor-ligand complex, the inactive conformation relative to the active conformations of an enzyme, and so on.

The equilibrium, $A \rightleftharpoons B$, is characterized by an equilibrium constant

$$K_{\text{eq}} = e^{-\Delta G/RT} \quad (1)$$

where ΔG is the free-energy difference between the two states, R is the gas constant, and T is the absolute temperature. Evolution acts on this equilibrium by favoring sequences that increase the fractional occupancy of B, the desirable state. The fractional occupancy, f , is defined as

$$f = \frac{[B]}{[A] + [B]} \quad (2)$$

Fig. 1. Overview of the design algorithm. (a) Fractional occupancy shows a sigmoidal dependence on ΔG , the equilibrium free-energy change [Eq. (4)]. ΔG_{offset} is the free-energy offset at which the target state, for example, the bound receptor, is half occupied. The plot illustrates three regimes: left, high occupancy (95–100%) and low sensitivity, where the fractional occupancy remains largely unaffected by mutations that induce moderate changes in ΔG ; middle, medium occupancy (5–95%) and high sensitivity, where there is significant change in the fractional occupancy due to changes in ΔG ; and right, low occupancy (0–5%) and low sensitivity, where the fractional occupancy is negligible and remains thus despite moderate improvements in ΔG . (b) Venn diagram graphically demonstrates the Boolean operators AND, OR, and NOT. AND and OR gates act on two states X and Y, whereas the NOT gate acts on one state, X. All states are drawn as circles and the selected area is marked in stripes. (c) Boolean operators (NOT, AND, and OR) are used to integrate the variables into an optimization objective function. By defining the three fundamental logical operators, we can formulate arguments encoding combinations of physical properties, including where properties tradeoff with one another. (d) The design flow chart exemplified for designing for stability and ligand binding, in four steps: first, the relevant properties of the system (stability of individual components and affinities between components) are defined. Second, for each property, the steepness and offset are defined and the sigmoidal function is set [Eq. (6)]. Third, the objective function is calculated by integrating the sigmoidal functions using Boolean operators. Fourth, the objective function is used by simulated annealing Monte Carlo sequence optimization. (e) Fitness is a complex function of biomolecular properties (such as binding, stability, etc.) exemplified here as a function of sequence exchanges at two positions only; in reality, fitness is a function in n -dimensional space, where n is the number of mutable protein or RNA residues. The landscape's ruggedness is the outcome of the nonadditive fitness effects of mutations, the limited choice of natural amino acids, and of tradeoffs between different molecular properties.

where [A] and [B] are the concentrations of the two states at chemical equilibrium [39]. Combining Eqs. (1) and (2) and the thermodynamics definition

$$K_{\text{eq}} = \frac{[B]}{[A]} \quad (3)$$

provides that the fractional occupancy of the desirable molecular state as a function of free-energy change of the process exhibits a sigmoidal relationship:

$$f = \frac{1}{1 + e^{\Delta G/RT}} \quad (4)$$

More generally, we rephrase Eq. (4) by adding an offset free-energy term ΔG_{offset} , at which the target state is half occupied (i.e., where [A] = [B]; Fig. 1a):

$$f = \frac{1}{1 + e^{(\Delta G - \Delta G_{\text{offset}})/RT}} \quad (5)$$

ΔG_{offset} reflects the energy requirements of a particular characteristic, namely, how favorable should the optimized trait's energy ΔG be to achieve a target occupancy (e.g., to obtain $f > 0.5$, ΔG must be smaller than ΔG_{offset}). For example, the receptor-ligand binding energy required to achieve half-maximal receptor occupancy depends on the ligand concentration; the lower the ligand concentration is, the higher ΔG_{offset} is, as well as the more favorable the required binding affinity must be to achieve a certain degree of occupancy [39,40]. Equation (5) asymptotically approaches 1 at negative ΔG values and 0 at positive ΔG values with an inflection point at $\Delta G = \Delta G_{\text{offset}}$ (Fig. 1a).

We further rephrase the sigmoidal relationship of Eq. (5) as:

$$f_x = \frac{1}{1 + e^{(x-o)s}} \quad (6)$$

where x and o are ΔG and ΔG_{offset} of Eq. (5), respectively, and s determines the slope of the sigmoidal response to free-energy changes around the inflection point.

Equation (6) generalizes the chemical equilibrium description of the fractional occupancy into the concept of biomolecular "fitness", reflecting the molecule's activity or functionality within an organism. The actual functionality of a biomolecule could have a sharper or a more attenuated dependence on the equilibrium free-energy change than Eq. (5) dictates. As an example, a molecule that has irreversible effects on cell fate, such as a nuclease, a protease, or a gene master regulator, may have large biological effects even at low fractional occupancies; the biomolecular output in this scenario is reproduced by low o and high s values. Conversely, two competing effector molecules, which bind overlapping sites on their target molecule, may only exert their biological effects at high occupancy [41] and would be modeled with high o and low s values (guidelines for determining o and s values are provided in the supplement). For simplicity, in what follows, we refer to fractional occupancy and biological "fitness" interchangeably, noting that both concepts could be modeled within the suggested framework.

The fractional occupancy of the desired state is certainly but one of many factors affecting the molecule's contribution to organism fitness, and expression levels, half-lives, post-translational modifications, and compartmentalization, play important complementary roles [42–45]. The sigmoidal relationship of Eqs. (5) and (6) is nevertheless a fundamental property of biomolecular systems, which can quantitatively describe numerous observations in natural and laboratory evolution, including the response to mutations [46]. Consider a concrete example of protein stability in mesophilic organisms. Many proteins have $\Delta G_{\text{folding}}$ values of -5 kcal/mol [47] meaning that, under standard conditions, more than 99.97% of the protein is in its folded state [Eq. (4)]. If protein misfolding [10], quality control [48], expression levels, post-translational modifications, and other extrinsic factors [49] are held constant, a mutation that decreases $\Delta G_{\text{folding}}$ by 2 kcal/mol would increase the folded fraction by less than 0.03% and would not be subject to strong evolutionary selection; thus, the high fractional occupancy regime exhibits relatively low sensitivity to mutation (Fig. 1a). By contrast, a protein with $\Delta G_{\text{folding}}$ of -1 kcal/mol is 84% folded, and a mutation that decreases $\Delta G_{\text{folding}}$ by 2 kcal/mol would increase the folded fraction by an appreciable 13% and would be strongly favored by selection, leading to high sensitivity to mutation around the sigmoid inflection point (Fig. 1a). Protein engineering is routinely used to successfully thermostabilize natural proteins [19,20,50,51], demonstrating that protein stability plateaus in evolution and that natural molecules can be readily engineered for higher stability. For other examples where physical traits plateau in evolution, see Supplemental Data.

A fuzzy-logic model for the selection of biomolecules with multiple physical traits

From the considerations mentioned above, we infer that a general model for representing multiple constraints selecting a biomolecule should exhibit two properties: first, a biomolecule's contribution to organism fitness is not only a function of the free-energy change associated with a physical trait but also a function of the fractional occupancy of the desirable state (Fig. 1a); and second, the constraints operating on the evolution of intrinsic physical traits are a complex, nonadditive function of the relevant target-state occupancies. To model the combination of constraints that shape a biomolecule through evolution, we borrow a concept from “fuzzy” or many-valued logic systems [52,53]. Fuzzy-logic formulations have been applied to produce graded rather than all-or-none responses to changes in input. It extends conventional Boolean logic by proposing that truth (1) and falsehood (0) are only extremes in a continuum of evaluations (0–1) with states that have intermediate values of truth and falsehood. The combination of molecular constraints, based on the fractional occupancies described above, lends itself naturally to fuzzy-logic treatment, in which the fractional occupancy f [Eq. (6)] of any energy term is treated as a Boolean variable and the standard Boolean operators (NOT, AND, and OR) are used to combine the variables into an optimization objective function (Fig. 1b): for any two target states (e.g., folded molecule, bound receptor, etc.), the individual fractional occupancies, f_x and f_y , are computed using Eq. (6). We thus define, as in Boolean algebra, $\neg f_x = 1 - f_x$ (NOT f_x), $f_x \wedge f_y = f_x * f_y$ (f_x AND f_y), and $f_x \vee f_y = f_x + f_y - f_x * f_y$ (f_x OR f_y) (Fig. 1b). In fuzzy logic, values approaching 1 satisfy the constraints embodied in the argument and values approaching 0 fail to satisfy them (Fig. 1c). By defining the three fundamental logical operators (Boolean NOT, AND, and OR), any argument comprising any number of constraints relating to energy evaluations of biomolecular states can be formulated in a statement that resembles natural language. The fuzzy-logic arguments then serve as objective functions (Fig. 1d) in stochastic sequence optimization in search for sequences that best conform to the encoded constraints (Fig. 1e). As in a natural language, the objective functions are an approximate model or representation of reality, and a completely accurate description of the fitness landscape, accounting for all of the constraints bearing on the biomolecular system, is impossible.

Despite this caveat, the fuzzy-logic system provides a framework for testing our understanding of the main constraints observed in biomolecular systems by integrating the fractional occupancies of Eq. (6) for each of the optimized energy criteria (e.g., binding and stability) using Boolean operators (Fig. 1c). In the supplement, we demonstrate that, under certain conditions, multistate design [30], which optimizes the energy gap between a target state and its competitors, describes a special case of the sigmoid relationship, assuming one target and several off-target states. Optimizing the energy gap suffers from a weakness, though, since mutations are favored even if they lead to the unwanted destabilization of the designed molecule or of the desired complex. In the context of the discussion mentioned above, energy-gap optimization, when applied with no additional constraints, fails to take into account that biomolecules need to stably fold in order to function correctly. Indeed, previous studies augmented energy-gap optimization with cutoffs that prevent energy terms from deteriorating beyond a certain limit [54,55]. Setting strict cutoffs on energy values is, however, more likely to funnel sampling to regions of sequence space where physical properties are on the border between optimality and suboptimality, potentially resulting in nonfunctional molecules. Using sigmoidal constraints and tradeoffs rather than strict cutoffs directs the search toward compromises that raise the chances of satisfying each of the individual constraints at least to sufficiency.

The fuzzy-logic design framework also allows framing constraints other than of the energy gap between states. For instance, several successful binder design studies selected sequences based on whether they encode both active-site side-chain rigidity and high-affinity binding [34–38]. Applying the fuzzy-logic strategy, we can explicitly optimize these properties during design simulations using the objective function:

$$O = B(a, b) \wedge S(a) \quad (7)$$

(Fig. 1c), where a and b are the designed protein and the target molecule, respectively, and B and S are the sigmoidal functions [Eq. (6)] for binding energy and stability, respectively. Indeed, in another paper, we demonstrated that, by optimizing Eq. (7), during antibody-design simulations, one can select antibody backbone conformations and sequences that exhibit features seen in natural antibody binders, such as conserved backbone conformations, hydrogen-bonding patterns, and van der Waals packing arrangements observed both within and between antibodies and their target molecules (Lapidot et al., unpublished results). Encouragingly, by optimizing Eq. (7), side-chain conformation stability at the designed binding surface was comparable and exceeded that observed in natural antibodies, even though, unlike previous design of function studies [37], side-chain rigidity was not explicitly used to select models. These results suggest that fuzzy-logic optimization could address some of the challenges of *de novo* design of function.

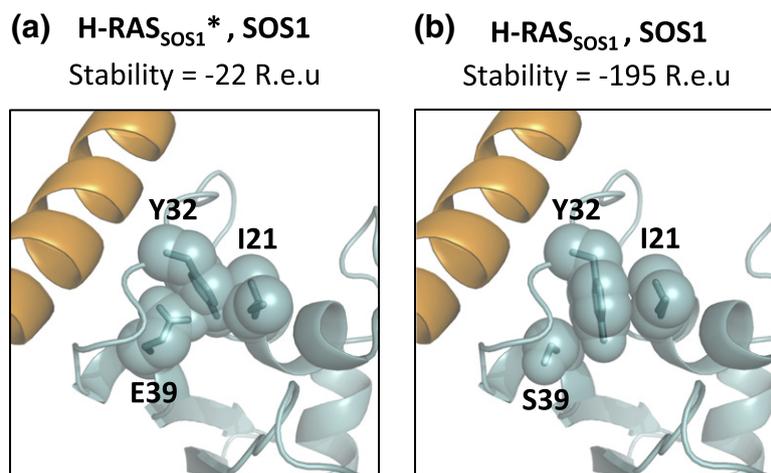


Fig. 2. Single-state, positive-only design violates the requirement for stability in alternative conformations. The single-state designed sequence of H-RAS_{PI3K} (H-RAS designed to target PI3K) was threaded on H-RAS_{SOS1} backbone (H-RAS in the SOS-1 bound backbone, cyan), producing the H-RAS_{SOS1}^{*},SOS1 structure (SOS1 is in orange) (a) after side-chain and rigid-body minimization. This designed sequence compromises the stability of H-RAS_{SOS1} monomer (b) by +173 R.e.u. mainly due to steric overlap among Y32, E39, and I21.

Here, we focus on additional uses for the fuzzy-logic framework and specifically on cases where tradeoffs between different molecular properties are prominent, such as binding specificity. We note, however, that this framework can be applied to situations where optimizing one property does not come at the expense of other properties and to situations where maximizing one trait is deleterious (as such in Ref. [29]). In the latter case, two sigmoid functions [Eq. (6)] can be applied to constrain the same physical trait, for example, $S_1(a)$ and $S_2(a)$, where S_1 defines the lower energy bound (with a low σ parameter) and S_2 defines the upper bound (high σ parameter) of the same energy criterion. The optimization objective function is then defined as:

$$O = S_2(a) \wedge \neg S_1(a) \quad (8)$$

biasing sequence search toward values intermediate between the two extremes.

Empirical tests of the fuzzy-logic design language

We subject the fuzzy-logic framework to two classic tests of computational protein design: the design of multispecificity and of orthogonal specificity. These tasks previously required different heuristics and optimization frameworks [30–33,54–60], but here we show that the fuzzy-logic language provides a unified framework for analyzing both phenomena.

Multispecificity design

As a model for multispecificity design, we select the protein human H-RAS, which has been co-crystallized with five natural partners [61–64] and was previously subjected to multispecificity analysis [33]. RAS is a member of the family of small GTP-binding proteins and is involved in regulating signal-transduction processes leading to cell growth and differentiation by directly controlling multiple pathways [65,66]. H-RAS presents special challenges to single-state design because it interacts *via* overlapping surfaces with several proteins, some of which have different folds and molecular functions, and because the backbone conformation of H-RAS varies when bound to its different partners [61–64]; the designed H-RAS sequences must therefore be compatible with all of H-RAS's backbone conformations and encode favorable contacts with all of its partners. To demonstrate the shortcomings of single-state design, that is, design toward any one of the H-RAS-bound structures, we select for stochastic sequence optimization eight amino acid positions on the H-RAS binding surface, four in switch region I (Ile36, Glu37, Asp38, and Ser39), two in switch region II (Gln61 and Glu63), and two outside of these regions (Ile21 and Gln25). With the use of RosettaDesign (Experimental Procedures), single-state, positive-only design of H-RAS toward each of its partners separately leads to sequences that

increase H-RAS stability and its binding affinity for that particular target. However, single-state design clearly violates the requirements of stability in the alternative H-RAS conformations and binding affinity for the other partners (Table S1). In some cases, the energy violations are severe [more than 10 R.e.u. (Rosetta energy units); Table S1 and Fig. 2], suggesting that optimizing H-RAS to bind any one of its partners alone increases specificity (as has been observed in other systems [67]) and protein rigidity, at the cost of binding to its other natural partners.

To model the multiple physical requirements on H-RAS, we use the new framework to encode energy constraints on all of the observed H-RAS backbone conformations and on binding affinities for all of its partners:

$$O = \mathbf{B}(\text{H-RAS}_{\text{SOS1}^*}, \text{SOS1}) \wedge \mathbf{B}(\text{H-RAS}_{\text{PI3K}^*}, \text{PI3K}) \wedge \mathbf{B}(\text{H-RAS}_{\text{BRY-2RBD}^*}, \text{BRY-2RBD}) \wedge \mathbf{B}(\text{H-RAS}_{\text{Ral-GDS}^*}, \text{Ral-GDS}) \wedge \mathbf{B}(\text{H-RAS}_{\text{RAS-GAP}^*}, \text{RAS-GAP}) \wedge \mathbf{S}(\text{H-RAS}_{\text{SOS1}^*}) \wedge \mathbf{S}(\text{H-RAS}_{\text{PI3K}^*}) \wedge \mathbf{S}(\text{H-RAS}_{\text{BRY-2RBD}^*}) \wedge \mathbf{S}(\text{H-RAS}_{\text{Ral-GDS}^*}) \wedge \mathbf{S}(\text{H-RAS}_{\text{RAS-GAP}^*}) \quad (9)$$

where O is the optimization objective function, H-RAS* is the designed H-RAS protein, and B and S are sigmoidal functions for binding and stability, respectively (see [Experimental Procedures](#)). Subscripts in Eq. (9) denote which backbone conformation is used for modeling H-RAS; thus, H-RAS_{SOS1}* denotes the designed H-RAS sequence modeled on the SOS1-bound backbone conformation.

Each of the functions (B and S) of Eq. (9) requires parameterization of the offset (o) and slope (s) as defined in Eq. (6). The offset and slope parameters were determined according to the guidelines in the supplement. We conduct 4000 independent design simulations starting from two single-state designed complexes as representatives of two different H-RAS-bound complexes, H-RAS,SOS1 and H-RAS,Ral-GDS, whereby the eight selected H-RAS positions are optimized to bind to each partner separately. In each simulation, the H-RAS sequence is subjected to nondeterministic simulated annealing Monte Carlo optimization with Eq. (9) serving as the optimization objective function (Fig. 1d and [Experimental Procedures](#)), and the resulting design variants are ranked according to the objective function [Eq. (9)]. In both cases, the top-ranked designs have binding affinities and stabilities similar to those computed for the natural H-RAS structures. Further, the top-ranked designs show favorable H-RAS-ligand molecular contacts as observed in the natural H-RAS (Fig. 3).

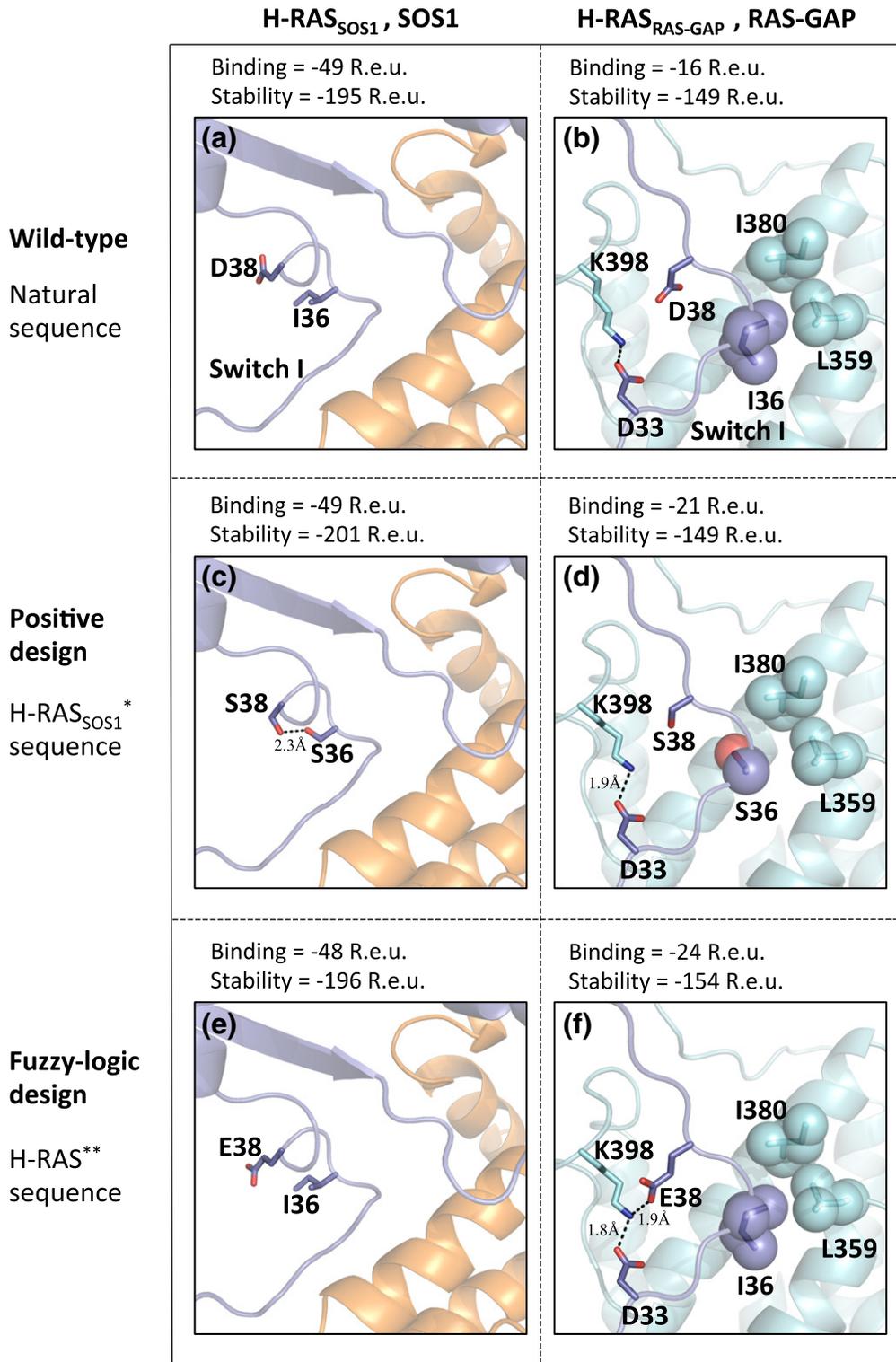
As expected, Monte Carlo sequence optimization does not converge on a single solution but rather produces a large landscape of sequences that are compatible with the constraints defined by the objective function. As the objective function of Eq. (9) encodes 10 different fractional occupancies (five each for binding and H-RAS stability), we set the threshold for favorable sequences at an objective value above $0.95^{10} = 0.59$, equivalent to requiring roughly 95% occupancy on all the terms of Eq. (9). For the case where we started with the H-RAS-SOS1 single-state designed model, 75% of the trajectories passed this threshold yielding approximately 1400 unique sequences. Notably, none of the H-RAS sequences optimized by single-state design to bind only SOS1 (H-RAS*,SOS1) (Fig. 3 and Table S1) overlapped any of the fuzzy-logic designed sequences. Further, the objective function calculated for the single-state designed H-RAS was 0.35, well below the acceptance threshold of 0.59, indicating that this design is suboptimal with respect to the set of constraints encoded in Eq. (9). Even though the resulting sequence space is large, on a position-by-position basis, the designed H-RAS proteins converge on amino acid identities that are physicochemically similar to the natural H-RAS (Fig. 4). Further, in several trajectories, produced from simulations that started with H-RAS designed solely to bind SOS1 and optimized using Eq. (9), the natural H-RAS sequence was retrieved as is and was ranked in the top 15% of all designed sequences. Although the natural H-RAS sequence is ranked high, there are 108 unique sequences in the set that converge to similar but not identical sequences (Fig. 4) and are predicted to have superior objective function values [Eq. (9)] compared to the natural H-RAS. Inaccuracies in the energy function and in the optimization objective function could account for this result, and other H-RAS properties that were not encoded by Eq. (9) might be compromised by these nonnatural sequences; alternatively, these solutions comprise variants predicted to have comparable or better stability and binding properties relative to the natural H-RAS for the chosen partners. Thus, the use of a gradual response curve [the sigmoid of Eq. (6) and Fig. 1a], rather than all-or-none thresholds, enables ranking the optimized sequences and provides testable hypotheses for structure–function studies.

Orthogonal-specificity design by combining positive and negative design constraints

The challenge in orthogonal-specificity design is generally defined as follows: starting from a natural complex of known structure, design a pair of proteins that bind as tightly as the natural complex but will bind their natural partners more weakly. Orthogonal specificity is a crucial determinant of the function of biological systems, insulating homologous signaling and metabolic pathways from one another, and has received several treatments from computational protein design [31,32,58,60,67].

To test the new design strategy in this context, we analyze two complexes from the bacterial E colicin endonuclease-immunity family, endonuclease-immunity 2 (E2,Im2) and endonuclease-immunity 9 (E9,Im9) [69].

The two pairs share the same overall folds, binding mode, and an identical set of residues at the core of the binding site (the interaction hotspot). There are, however, significant differences in the identities of interface residues outside the core. Additionally, one of the interfacial loops on the immunity protein differs in conformation, resulting in a 5.6° reorientation of the partners in one complex relative to the other [70]. The net effect of these molecular differences is that both cognate pairs are highly specific and bind with femtomolar dissociation constants, while the noncognate pairs interact with 6–7 orders of magnitude lower affinities [71,72].



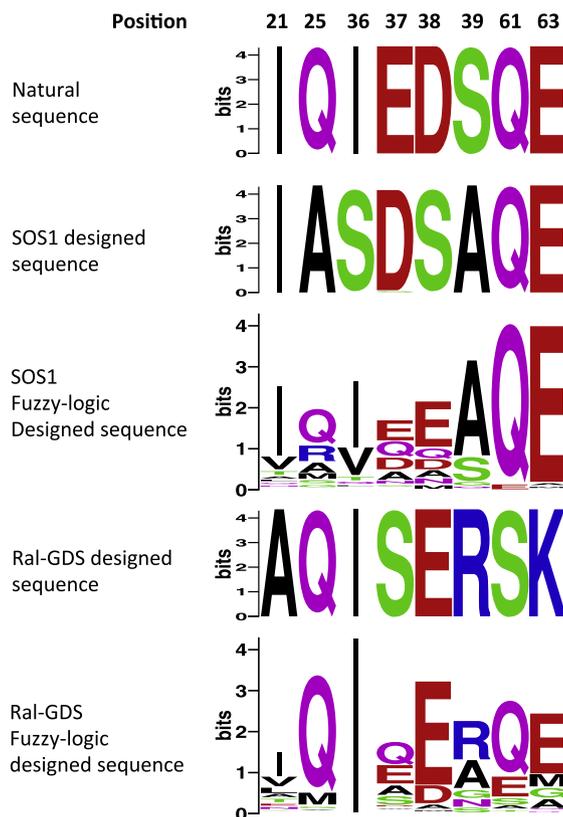


Fig. 4. Fuzzy-logic design converges on amino acids with physicochemical properties that resemble those of the natural H-RAS. Sequence logos [68] of the natural, positive-only, and fuzzy-logic designed sequences of two structures with distinct conformations: H-RAS_{SOS1} and H-RAS_{Ral-GDS}. All designable positions exhibit two features: (i) a different sequence is compatible with each structure after positive design and (ii) the sequences of the two converge after the design to identities physicochemically similar to those observed in the natural sequence.

This colicin family has been subjected previously to a specialized heuristic for specificity design producing designed partners that bound one another 2–3 orders of magnitude more tightly than their interactions with the noncognate natural partners and with 2–4 orders of magnitude lower affinity than observed for natural colicin pairs [56,58].

Here, we show that subjecting the E9-Im9 naturally occurring high-affinity complex to the orthogonal-specificity design objective function defined below reproduces the required energy and structure characteristics seen in natural high-specificity pairs and in more specialized computational specificity-design heuristics:

$$O = \mathbf{B}(E9^*, Im9^*) \wedge \mathbf{S}(E9^*) \wedge \mathbf{S}(Im9^*) \wedge \neg \mathbf{B}(E9^*, Im9) \wedge \neg \mathbf{B}(E9, Im9^*) \quad (10)$$

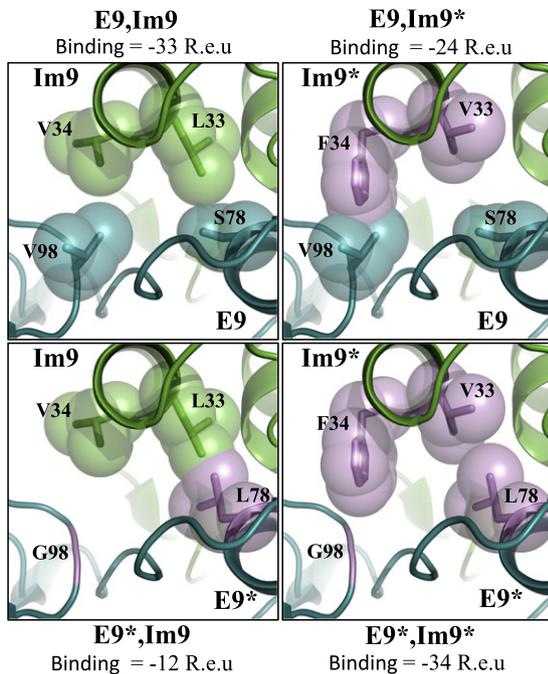
where B and S are the binding and stability fractional occupancies determined from Eq. (6), E9 and Im9 are the natural sequences, and E9* and Im9* are the designed sequences. The first three terms in Eq. (10) ensure that the designed complex affinity and the designed monomer stabilities are maintained close to their levels in the natural complex, while the last two terms disfavor cross-binding of each of the designed partners to its original natural

Fig. 3. Single-state design *versus* fuzzy-logic design of the multispecific hub protein H-RAS. This view focuses on H-RAS switch I loop. The two H-RAS complexes presented here, H-RAS_{SOS1} (a–e) and H-RAS_{RAS-GAP} (b–f), have different backbone conformations. Upper panel: structures of the native complexes of H-RAS_{SOS1} (violet) and SOS1 (orange) complex (a) and H-RAS_{RAS-GAP} (violet) and RAS-GAP (cyan) complex (b), where I380, L359, and I36 form favorable van der Waals contacts and K398 and D33 form a salt bridge. Middle panel: single-state design of H-RAS_{SOS1}*-SOS1 complex, where D38S and I36S are introduced to form a hydrogen bond (c), and the designed sequence of H-RAS_{SOS1}* threaded on H-RAS_{RAS-GAP} backbone, where I36S packing with I380 and L359 is compromised (d). Bottom panel: fuzzy-logic design of H-RAS in complex with SOS1 (e) and with RAS-GAP (f). Position 36 was restored to Ile, which is the natural identity, while position 38 was designed to Glu, which forms a new electrostatic interaction with K398 in H-RAS_{RAS-GAP}*,RAS-GAP structure (f), resulting in a sequence that is similar to the wild type.

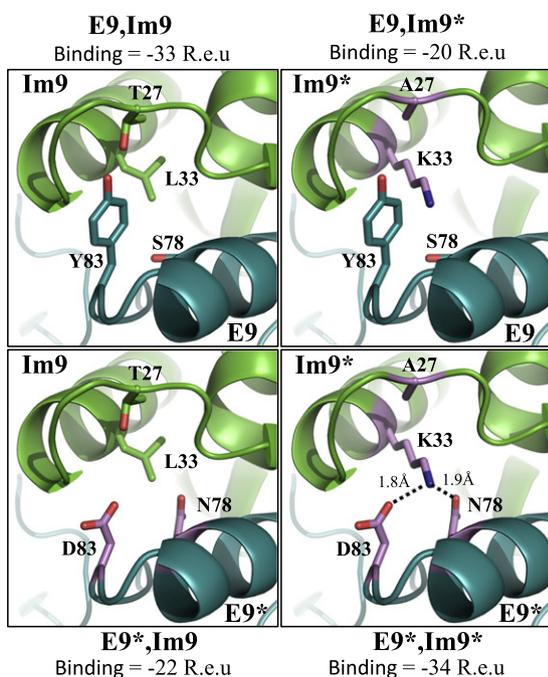
partner. Thus, Eq. (10) demonstrates the flexibility and expressiveness of the fuzzy-logic framework, which can be readily extended to situations where some target states are disfavored.

The colicin pairs 2 and 9 present a unique case, where molecular structures for two cognate pairs [70,73] and one of the noncognate pairs (E9, Im2) [74] are available. These structures allow us to estimate the computed energies that are required for establishing high-affinity cognate binding and medium-affinity noncognate binding in our simulations. The two cognate pairs (E2,Im2 and E9,Im9) have calculated binding affinities below -32 R.e.u., whereas the noncognate pair (E9,Im2) has calculated binding affinity of -27 R.e.u. We therefore set the offset value for noncognate binding [ρ in Eq. (6)] to -27 R.e.u., meaning that

(a) Knobs-into-holes:



(b) Polarity switch:



the fractional occupancy value [Eq. (6)] for a designed noncognate pair with computed binding energy as the natural noncognate E9,Im2 pair would be one-half (all other parameters are provided in Supplemental Experimental Procedures).

Starting from the molecular structure of the E9-Im9 complex [73], we select nine positions at the binding surfaces of the immunity and endonuclease pair that were experimentally shown to contribute to specificity [four on the immunity (Thr27, Leu33, Val34, Thr38) and five on the endonuclease (Ser77, Ser78, Tyr83, Lys97, Val98)] [71]. These positions are subjected to stochastic sequence optimization using the objective function defined in Eq. (10). Unlike the specialized schemes used previously [58], we employ the same simulated annealing Monte Carlo sampling method used above for the calculation of multispecificity. The difference between the H-RAS multispecificity implementation above and colicin orthogonal-specificity design here is

Fig. 5. Two designed specificity switches in the colicin endonuclease-immunity pair E9-Im9. Interactions with binding energy greater than or equal to -27 R.e.u. are noncognate (see Results), while interactions with binding energy less than or equal to -33 are expected to be as favorable as the natural pairs. Im9 and E9 are shown in green and deep-teal cartoons, respectively. The mutated residues in the design pair are shown in violet. (a) A design pair in which specificity is encoded by knobs-into-holes features. In the natural pair (upper left panel), L33 and V34 of Im9 pack against S78 and V98 of E9, respectively. In the design pair (lower right panel), L33V mutation introduces a smaller side chain, and the complementary S78L mutation maintains packing density as in the natural pair. Similarly, V34F introduces a large side-chain in the designed Im9* while V98G removes a side-chain. In the designed noncognate pair E9,Im9*, V34F forms a steric overlap with E9's V98, thus destabilizing the bound state. Similarly, in the designed noncognate pair E9*, Im9 mutation S78L forms a steric overlap with Im9's L33. (b) A design pair, in which specificity is encoded through a switch in polarity features. The natural interaction is mainly based on hydrophobic packing. In the designed pair (bottom right), K33 forms hydrogen bonds with D83 and N78, as well as a salt bridge with D83. In the designed noncognate pairs, the hydrogen bonds are not formed and polar residues pack against hydrophobic ones leading to electrostatic frustration similar to the observation in the natural noncognate pair [74]. Also, there are steric overlaps in the designed noncognate pairs that further destabilize the bound state. In E9*, Im9's (bottom left) D83 and N78 pack against the hydrophobic L33. A steric overlap is formed between L33 and N78. In the design noncognate pair E9,Im9*, K33 packs against S78 and Y83, resulting in steric overlap.

restricted solely to the objective function [Eqs. (9) and (10), respectively].

As in the design of multispecificity mentioned above, orthogonal-specificity design does not converge on a single solution, but rather it produces hundreds of sequences that meet the threshold values for individual partner stabilities, high cognate binding affinity, and low affinities toward the respective natural (noncognate) partners (Fig. 5). In fact, for the top designs, the resulting energy gap between the designed cognate and designed noncognate pairs (> 10 R.e.u.; Fig. 5) is calculated to be larger than that computed for the natural pairs. The molecular details of the designed complexes reproduce the hallmarks of high-affinity and high-specificity complexes, as seen in natural colicin complexes [74], including intricate van der Waals packing and hydrogen-bonding networks in the designed cognate complexes (E9*,Im9*). Conversely, the modeled noncognate complexes (E9*,Im9 and E9,Im9*) exhibit electrostatic and steric frustration (Fig. 5). Even though we use only a single design heuristic and objective function, different trajectories result in quite different solutions to encoding specificity; these include, in the two cases highlighted in Fig. 5, one that employs knobs-into-holes to introduce steric overlaps in noncognate pairs and another that uses electrostatic frustration to achieve the same goal. Fuzzy-logic design can therefore define several physicochemically distinct ways of achieving the same desired function. We note, however, that, in the natural colicin family, specificity is also encoded by backbone conformation changes and perturbations to the rigid-body orientations [74] and that recapitulating natural sequences in this family likely requires modeling these degrees of freedom. The objective function defined in Eq. (10) can indeed be used to guide an optimization algorithm that samples sequence, as well as backbone and rigid-body conformations (Lapidoth et al., unpublished results) [56,75].

Discussion

The evolution of a biological macromolecule is subject to selection pressures that bear on its physical properties, such as stability, binding affinity, specificity, and activity. We presented a framework for encoding each physical constraint based on the equilibrium-thermodynamics concept of fractional occupancy of the target state within an ensemble of states. In the example of H-RAS, multispecificity design specified a large number of physical properties, overall encoding 10 different constraints relating to both stability and binding affinity. These 10 constraints were combined using Boolean operators to recapitulate sequences and conformations observed in nature, including the natural H-RAS sequence. By changing the objective function only and without changing the underlying optimization algorithm, we can also model competition with alternative states, as seen in specificity-switch modeling. The result is a general language for encoding different physical traits, including ones that exhibit tradeoffs with one another, as logical expressions that can be translated into a numerical objective function for optimization. In reality, the tradeoffs between different physical properties extend beyond purely thermodynamics properties and also involve kinetics [15,76] and, specifically, association and dissociation rate constants. Where structural data are available on kinetic intermediates and parameters, kinetic constraints could also be modeled using the same framework, as could other important properties such as selectivity. We expect the fuzzy-logic design scheme to work well in cases where large structural differences between the alternative states exist and where the desired energy gaps between target and off-target states are large, as in the case studies selected here.

Encoding fine discrimination between structurally similar states or small energy gaps may require refined energy functions.

Not all properties of biological systems are explained within the context of optimality and adaptation [77]; however, the physical constraints shaping biomolecules and biomolecular structures are largely known. The computational framework suggested here can be useful for a quantitative analysis of these constraints and for the design of novel biomolecules with desired functions. The design of novel or rewired signaling and metabolic networks [78], for instance, inevitably invokes complex tradeoffs between different molecules and their properties, and these can also be encoded within the framework presented here. Recent advances in the application of deep sequencing to libraries of natural protein variants [79,80] and of *in vitro* mutational repertoires selected for complex combinations of physical features, including stability, binding, and specificity profiles [34,36,81–83] are generating datasets comprising thousands of mutants that relate sequence changes to function at unprecedented resolution and coverage [84]. The fuzzy-logic design framework described here can be used to test hypotheses relating sequence, structure, and energetics to function, as well as in turn to fitness. Interpreting these empirical data through fuzzy-logic design may ultimately lead to a better understanding and ability to manipulate the multiple physical properties and tradeoffs that shape the evolution of biomolecular functions. More broadly, tradeoffs in biology extend beyond the molecular level to the evolution of organism phenotypes [85]; the use of sigmoid functions to represent the levels at which individual traits reach sufficiency and fuzzy logic to integrate these into optimization objective

functions that resemble natural language could thus be generalized beyond the molecular level explored here.

Experimental Procedures

Sequence optimization

For both multispecificity and orthogonal-specificity design, we start from wild-type complexes of solved molecular structure, relaxed using procedures described in Supplemental Experimental Procedures, and compute the binding and stability energies of each component in the natural system. For each design objective, the o and s parameters of the fractional occupancy equation [Eq. (6)] are computed to reflect the desired energies of each modeled state (see Results and Supplemental Experimental Procedures). Each design task is assigned a different objective function [e.g., Eqs. (9) and (10)] to reflect its unique biomolecular constraints. We then select a set of interface residues for stochastic sequence optimization. At each step of the optimization trajectory, a single position is chosen at random out of the set of positions allowed to optimize (including on both sides of the interface in the case of orthogonal-specificity design), and a substitution to a randomly chosen identity is introduced in all modeled protein states (i.e., unbound for stability calculations and bound to each of the partners in the case of binding calculations); each state is then subjected to the same energy-relaxation protocol, including combinatorial side-chain packing and side-chain and rigid-body minimization. All energies are computed with the Rosetta all-atom energy function, which is dominated by contributions from van der Waals packing, hydrogen bonding, and solvation [86]. A new state is accepted subject to the Metropolis criterion and the mutation and selection process is repeated. Additional implementation details are provided in Supplemental Experimental Procedures, including complete RosettaScripts [87] and execution instructions for each design task.

Binding and stability calculations

Binding energies are computed as the difference in system energy between the bound repacked model and the repacked monomers when taken apart from one another and are substituted for x in Eq. (6) to produce the bound fractional occupancy [B in Eqs. (7)–(10)]. Monomer stability is the system energy of the monomer in its bound conformation but dissociated from the ligand, and it is likewise substituted for x in Eq. (6) to produce the folded fractional occupancy [S in Eqs. (7)–(10)]. For example, $B(\text{H-RAS}_{\text{partner}^*}, \text{partner})$ of Eq. (9) is the fractional occupancy as a function of the binding energy of

the H-RAS mutant for its partner and $S(\text{H-RAS}_{\text{partner}^*})$ is the fractional occupancy as a function of the stability of the H-RAS mutant monomeric form modeled on the backbone observed in the partner-bound molecular structure.

Acknowledgments

We thank all members of the Fleishman laboratory who read and commented on this manuscript, as well as Ingemar Andre and Amnon Horovitz for critical discussion. Research in the Fleishman laboratory is supported by the Israel Science Foundation through an individual research grant and through the Center of Research Excellence (I-CORE) in Structural Cell Biology, the Human Frontier Science Program, a Marie Curie Reintegration Grant, a European Research Council Starter's Grant, an Alon Fellowship, the Yeda-Sela Center, the Geffen Fund, and a charitable donation from Sam Switzer. S.J.F. is the incumbent of the Martha S Sagon Career Development Chair.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2014.10.002>.

Received 30 July 2014;

Received in revised form 21 September 2014;

Accepted 2 October 2014

Available online 13 October 2014

Keywords:

stability-activity tradeoff;
Rosetta;
diminishing returns;
sequence optimization;
multistate design

References

- [1] Rooman MJ, Wodak SJ. Extracting information on folding from the amino acid sequence: consensus regions with preferred conformation in homologous proteins. *Biochemistry* 1992;31:10239–49.
- [2] Wolynes PG. Energy landscapes and solved protein-folding problems. *Philos Trans R Soc A Math Phys Eng Sci* 2005;363:453–64.
- [3] Li W, Wolynes PG, Takada S. Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proc Natl Acad Sci U S A* 2011;108:3504–9.
- [4] Sali A, Shakhnovich E, Karplus M. How does a protein fold? *Nature* 1994;369:248–51.
- [5] Kraur J. How do enzymes work? *Science* 1988;242:533–9.

- [6] Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht A a, Titchmarsh J, et al. Optimizing protein stability *in vivo*. *Mol Cell* 2009;36:861–71.
- [7] Beadle BM, Shoichet BK. Structural bases of stability–function tradeoffs in enzymes. *J Mol Biol* 2002;321:285–96.
- [8] Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH. Thermodynamic prediction of protein neutrality. *Proc Natl Acad Sci U S A* 2005;102:606–11.
- [9] Meiering EM, Serrano L, Fersht A R. Effect of active site residues in barnase on activity and stability. *J Mol Biol* 1992; 225:585–9.
- [10] Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 2009;19:596–604.
- [11] Tokuriki N, Stricher F, Serrano L, Tawfik DS. How protein stability and new functions trade off. *PLoS Comput Biol* 2008; 4:e1000002.
- [12] Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem* 2010;79:471–505.
- [13] Wang X, Minasov G, Shoichet BK. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 2002;320:85–95.
- [14] Zarrinpar A, Park S-H, Lim W a. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 2003;426:676–80.
- [15] Tawfik DS. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr Opin Chem Biol* 2014;21:73–80.
- [16] Mohammadi M, Olsen SK, Ibrahim Oa. Structural basis for fibroblast growth factor receptor activation. *Cytokine Growth Factor Rev* 2005;16:107–37.
- [17] Fleishman SJ, Baker D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* 2012;149: 262–73.
- [18] Bowie JU, Ltcy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 1990;253:164–70.
- [19] Korkegian A, Black ME, Baker D, Stoddard BL. Computational thermostabilization of an enzyme. *Science* 2005;308:857–60.
- [20] Miklos AE, Kluwe C, Der BS, Pai S, Sircar A, Hughes R a. Structure-based design of supercharged, highly thermoresistant antibodies. *Chem Biol* 2012;19:449–55.
- [21] Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003;302:1364–8.
- [22] Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, et al. Principles for designing ideal protein structures. *Nature* 2012;491:222–7.
- [23] Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, et al. Kemp elimination catalysts by computational enzyme design. *Nature* 2008;453:190–5.
- [24] Jiang L, Althoff E a, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, et al. *De novo* computational design of retroaldol enzymes. *Science* 2008;319:1387–91.
- [25] King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 2012;336: 1171–4.
- [26] Jha RK, Leaver-Fay A, Yin S, Wu Y, Butterfoss GL, Szyperski T, et al. Computational design of a PAK1 binding protein. *J Mol Biol* 2010;400:257–70.
- [27] Karanicolas J, Corn JE, Chen I, Joachimiak La, Dym O, Peck SH. A *de novo* protein binding pair by computational design and directed evolution. *Mol Cell* 2011;42:250–60.
- [28] Der BS, Edwards DR, Kuhlman B. Catalysis by a *de novo* zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* 2012; 51:3933–40.
- [29] Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun* 2012;3:1257.
- [30] Havranek JJ, Harbury PB. Automated design of specificity in molecular recognition. *Nat Struct Biol* 2003;10:45–52.
- [31] Grigoryan G, Reinke AW, Keating AE. Design of protein–interaction specificity gives selective bZIP-binding peptides. *Nature* 2009;458:859–64.
- [32] Ashworth J, Taylor GK, Havranek JJ, Quadri SA, Stoddard BL, Baker D. Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res* 2010;38:5601–8.
- [33] Humphris EL, Kortemme T. Design of multi-specificity in protein interfaces. *PLoS Comput Biol* 2007;3:1591–604.
- [34] Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman S, Su M, et al. Computational design of a protein-based enzyme inhibitor. *J Mol Biol* 2013;425:3563–75.
- [35] Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch E-M, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 2011;332:816–21.
- [36] Strauch E-M, Fleishman SJ, Baker D. Computational design of a pH-sensitive IgG binding protein. *Proc Natl Acad Sci U S A* 2014;111:675–80.
- [37] Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 2013;501:212–6.
- [38] Fleishman SJ, Khare SD, Koga N, Baker D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci* 2011;20:753–7.
- [39] Kuriyan J, Konforti B, Wemmer D. *The molecules of life*. New York: Garland Science; 2012.
- [40] Krohn K a, Link JM. Interpreting enzyme and receptor kinetics: keeping it simple, but not too simple. *Nucl Med Biol* 2003;30:819–26.
- [41] Hirschi A, Cecchini M, Steinhardt RC, Chamber MR, Dick FA, Rubin SM. An overlapping kinase and phosphatase docking site regulates activity of the retinoblastoma protein. *Nat Struct Mol Biol* 2010;17:1051–7.
- [42] Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* 2005;436:588–92.
- [43] Beltrao P, Bork P, Krogan NJ, van Noort V. Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol* 2013;9:714.
- [44] Diekmann Y, Pereira-Leal JB. Evolution of intracellular compartmentalization. *Biochem J* 2013;449:319–31.
- [45] Ciechanover A. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* 2005;6:79–87.
- [46] Li W, Keeble AH, Giffard C, James R, Moore GR, Kleanthous C. Highly discriminating protein–protein interaction specificities in the context of a conserved binding energy hotspot. *J Mol Biol* 2004;337:743–59.
- [47] Fersht AR. *Structure and mechanism in protein science*. New York: Freeman; 1999.
- [48] Bershtein S, Mu W, Serohijos AWR, Zhou J, Shakhnovich EI. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol Cell* 2013; 49:133–44.
- [49] Serohijos AW, Shakhnovich EI. Merging molecular mechanism and evolution: theory and computation at the interface of

- biophysics and evolutionary population genetics. *Curr Opin Struct Biol* 2014;26:84–91.
- [50] Scott DJ, Kummer L, Tremmel D, Plückthun A. Stabilizing membrane proteins through protein engineering. *Curr Opin Chem Biol* 2013;17:427–35.
- [51] Wörn A, Plückthun A. Stability engineering of antibody single-chain Fv fragments. *J Mol Biol* 2001;305:989–1010.
- [52] Zadeh LA. Fuzzy sets. *Inf Control* 1965;8:338–53.
- [53] Hassanien AE, Al-Shammari ET, Ghali NI. Computational intelligence techniques in bioinformatics. *Comput Biol Chem* 2013;47:37–47.
- [54] Leaver-Fay A, Jacak R, Stranges PB, Kuhlman B. A generic program for multistate protein design. *PLoS One* 2011;6:e20937.
- [55] Allen BD, Mayo SL. An efficient algorithm for multistate protein design based on FASTER. *J Comput Chem* 2009;31:904–16.
- [56] Joachimiak LA, Kortemme T, Stoddard BL, Baker D. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein–protein interface. *J Mol Biol* 2006;361:195–208.
- [57] Sammond DW, Eletr ZM, Purbeck C, Kuhlman B. Computational design of second-site suppressor mutations at protein–protein interfaces. *Proteins* 2010;78:1055–65.
- [58] Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein–protein interaction specificity. *Nat Struct Mol Biol* 2004;11:371–9.
- [59] Yanover C, Fromer M, Shifman JM. Dead-end elimination for multistate protein design. *J Comput Chem* 2007;28:2122–9.
- [60] Havranek JJ. Specificity in computational protein design. *J Biol Chem* 2010;285:31095–9.
- [61] Pacold ME, Suire S, Perisic O, Lara-Gonzalez S, Davis CT, Walker EH, et al. Crystal structure and functional analysis of Ras binding to its effector phosphoinositide 3-kinase gamma. *Cell* 2000;103:931–43.
- [62] Scheffzek K. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* 1997;277:333–8.
- [63] Boriack-sjodin PA, Margarit SM, Bar-sagi D, Kuriyan J. The structural basis of the activation of Ras by Sos. *Nature* 1998;394:337–43.
- [64] Scheffzek K, Grünewald P, Wohlgemuth S, Kabsch W, Tu H, Wigler M, et al. The Ras-Byr2RBD complex: structural basis for Ras effector recognition in yeast. *Structure* 2001;9:1043–50.
- [65] Rosseland CM, Wierød L, Flinder LI, Oksvold MP, Skarpen E, Huitfeldt HS. Distinct functions of H-Ras and K-Ras in proliferation and survival of primary hepatocytes due to selective activation of ERK and PI3K. *J Cell Physiol* 2008;215:818–26.
- [66] Xue L. The Ras/phosphatidylinositol 3-kinase and Ras/ERK pathways function as independent survival modules each of which inhibits a distinct apoptotic signaling pathway in sympathetic neurons. *J Biol Chem* 2000;275:8817–24.
- [67] Shifman JM, Mayo SL. Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* 2002;323:417–23.
- [68] Crooks GE, Hon G, Chandonia J, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–90.
- [69] Papadakos G, Wojdyla Ja, Kleanthous C. Nuclease colicins and their immunity proteins. *Q Rev Biophys* 2012;45:57–103.
- [70] Wojdyla JA, Fleishman SJ, Baker D, Kleanthous C. Structure of the ultra-high-affinity colicin E2 DNase-Im2 complex. *J Mol Biol* 2012;417:79–94.
- [71] Keeble AHA, Joachimiak LL a, Maté MMJ, Meenan N, Kirkpatrick N, Baker D, et al. Experimental and computational analyses of the energetic basis for dual recognition of immunity proteins by colicin endonucleases. *J Mol Biol* 2008;379:745–59.
- [72] Keeble AH, Kirkpatrick N, Shimizu S, Kleanthous C. Calorimetric dissection of colicin DNase-immunity protein complex specificity. *Biochemistry* 2006;45:3243–54.
- [73] Kühlmann UC, Pommer a J, Moore GR, James R, Kleanthous C. Specificity in protein–protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J Mol Biol* 2000;301:1163–78.
- [74] Meenan N a G, Sharma A, Fleishman SJ, Macdonald CJ, Morel B, Boetzel R, et al. The structural and energetic basis for high selectivity in a high-affinity protein–protein interaction. *Proc Natl Acad Sci USA* 2010;107:10080–5.
- [75] Ollikainen N, Smith C a, Fraser JS, Kortemme T. Flexible backbone sampling methods to model and design protein alternative conformations. vol. 523. 1st ed. Elsevier Inc; 2013.
- [76] Knowling S, Bartlett AI, Radford SE. Dissecting key residues in folding and stability of the bacterial immunity protein 7. *Protein Eng Des Sel* 2011;24:517–23.
- [77] Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc B Biol Sci* 1979;205:581–98.
- [78] Dueber JE, Yeh BJ, Bhattacharyya RP, Lim Wa. Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. *Curr Opin Struct Biol* 2004;14:690–9.
- [79] Rhee S-Y, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 2003;31:298–303.
- [80] Fisher R, van Zyl GU, Travers S a a, Kosakovsky Pond SL, Engelbrech S, Murrell B, et al. Deep sequencing reveals minor protease resistance mutations in patients failing a protease inhibitor regimen. *J Virol* 2012;86:6231–7.
- [81] Whitehead T a, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 2012;30:543–8.
- [82] Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence–function relationships. *Nat Methods* 2010;7:741–6.
- [83] Pál G, Kouadio J-LK, Artis DR, Kossiakoff A a, Sidhu SS. Comprehensive and quantitative mapping of energy landscapes for protein–protein interactions by rapid combinatorial scanning. *J Biol Chem* 2006;281:22378–85.
- [84] Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. *Nat Methods* 2014;11:801–7.
- [85] Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo a. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* 2012;336:1157–60.
- [86] Das R, Baker D. Macromolecular modeling with rosetta. *Annu Rev Biochem* 2008;77:363–82.
- [87] Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* 2011;6:e20161.