

Searching for the Pareto frontier in multi-objective protein design

Vikas Nanda^{1,2}  · Sandeep V. Belure^{1,2} · Ofer M. Shir^{3,4}

Received: 27 May 2017 / Accepted: 25 July 2017 / Published online: 10 August 2017
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag GmbH Germany 2017

Abstract The goal of protein engineering and design is to identify sequences that adopt three-dimensional structures of desired function. Often, this is treated as a single-objective optimization problem, identifying the sequence–structure solution with the lowest computed free energy of folding. However, many design problems are multi-state, multi-specificity, or otherwise require concurrent optimization of multiple objectives. There may be tradeoffs among objectives, where improving one feature requires compromising another. The challenge lies in determining solutions that are part of the Pareto optimal set—designs where no further improvement can be achieved in any of the objectives without degrading one of the others. Pareto optimality problems are found in all areas of study, from economics to engineering to biology, and computational methods have been developed specifically to identify the Pareto frontier. We review progress in multi-

objective protein design, the development of Pareto optimization methods, and present a specific case study using multi-objective optimization methods to model the tradeoff between three parameters, stability, specificity, and complexity, of a set of interacting synthetic collagen peptides.

Keywords Computational protein design · Pareto optimality · Multi-objective optimization · Peptide · Collagen · Interactome

Introduction

The ability to engineer proteins or design them from scratch holds the promise for the control of chemical and spatial properties of materials at the nanoscale. Despite a limited alphabet of 20 amino acids, living organisms have evolved proteins as highly efficient catalysts, materials with photonic properties, molecular nanomachines capable of converting between light, chemical, and mechanical energy, and sophisticated signaling and regulatory networks. Still, natural evolutionary processes have accessed only an infinitesimal slice of potential protein sequences (Povolotskaya and Kondrashov 2010), suggesting an immense expanse of undiscovered protein functionality waiting to be discovered.

The field is making forays into this unexplored territory, developing proteins as therapeutics to fight disease and infection, as materials for biomedical applications, and as catalysts for industrial or household use (Braxton and Wells 1992; Jiang et al. 2008; Röthlisberger et al. 2008; Whitehead et al. 2012). The computer has been an essential tool in these explorations, employing physical or statistical models of sequence–structure–function relationships in a high-throughput fashion, sampling large swaths of sequence space to identify

This article is part of a Special Issue on ‘IUPAB Edinburgh Congress’ edited by Damien Hall.

✉ Vikas Nanda
nandavi@cabm.rutgers.edu

¹ Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ, USA

² Department of Biochemistry and Molecular Biophysics, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ, USA

³ Department of Computer Science, Tel-Hai College, Kiryat Shmona, Upper Galilee, Israel

⁴ The Galilee Research Institute—Migal, Kiryat Shmona, Upper Galilee, Israel

promising candidates for characterization and development in the laboratory. The use of computational protein design is rapidly expanding, pushing the state-of-the-art in computational models that efficiently and accurately compute the quality of candidate sequences, and algorithmic or hardware advances that promote rapid sampling of sequence space.

To apply computational methods to a specific design problem, it is necessary to have a defined computable objective. In nearly all projects, the central objective is optimizing the thermodynamic stability of a unique folded state that is able to perform the desired function. Thermodynamic stability is computed using chemical models of various degrees of resolution from heuristic sequence-based scoring functions (Nautiyal et al. 1995; Summa et al. 2002) to high-accuracy but computationally expensive quantum mechanics calculations of energetics (Kiss et al. 2013). The majority of computational protein design platforms calculate energetics of interactions at the atomic level, emphasizing non-bonding interactions, i.e., van der Waals packing, hydrogen bonding, and electrostatics (Kuhlman et al. 2003). However, other objectives may be desirable: solubility, immunogenicity, toxicity, cell permeability, dynamics, or functional efficiency. Each requires an appropriate quantitative model relating the objective to protein sequence for it to be computationally optimizable. Tools for evaluating solubility (Sormanni et al. 2015) can be applied to optimize protein libraries for solubility and enhance the concentration and shelf life of protein formulations. Humanization of therapeutic antibodies using tools such as those by Choi et al. (2015) and Griswold and Bailey-Kellogg (2016) can be applied to minimize adverse immune responses during administration. At the end of this review, we discuss in some detail the specific problem of designing complex sets of specifically interacting molecules, focusing on collagen mimetic peptide studies from our laboratory.

There are a number of well-developed global search heuristics for addressing single-objective optimization problems. Effective approaches should converge on the same set of globally optimal solutions regardless of the starting point, although in practice, the degeneracy of solutions for a particular fold can be large, and, often, the practical goal is to identify solutions that are “good enough” instead of globally optimal. Simulated annealing (Hellings and Richards 1994) and genetic algorithms (Voigt et al. 2000) are commonly used to sample high-barrier paths along a sequence trajectory between the starting point and a target solution.

Multi-objective protein design

In practice, protein design requires consideration of multiple objectives. This can complicate optimization if the quantitative measures of the multiple objectives are in different units, i.e., how would the stability of a therapeutic protein, measured in terms of free energy of folding, be compared with toxicity

based on the maximum tolerated dose? Furthermore, objectives may be in conflict, resulting in a tradeoff where optimizing one compromises the other. For example, the objective of optimizing the stability of a target protein fold may be achieved by maximizing the number of favorable residue–residue interactions throughout the structure. However, this does not take into account the stability of that sequence in competing compact states that differ from the native fold, i.e., the specificity of a sequence for the unique target state. It was shown using lattice chain models of simple protein heteropolymers that the stability and specificity of folding are in fundamental conflict (Chan and Dill 1991). Introducing hydrophobic residues into a sequence will stabilize the native fold, but also increase the degeneracy of states it is likely to adopt (Handel et al. 1993). The conflict between stability and specificity also exists in protein–protein binding interactions. The signaling protein calmodulin interacts with hundreds of targets and achieves this multi-specificity at the expense of the stability of individual calmodulin–target interactions (Shifman and Mayo 2002; Fromer and Shifman 2009).

Stability is also often at odds with other desirable design objectives. Increasing hydrophobicity may affect solubility, expression yield, and complicate purification. In designing enzymes, enhancing stability may dampen protein dynamics that facilitate catalysis, a phenomenon seen in nature where thermophilic versions of enzymes are inactive at room temperature (Howell et al. 2014). Enhancing the stability of therapeutic proteins may improve pharmacokinetic properties such as in vivo half-life (Hall 2014), but may also increase their immunogenicity (Camacho et al. 2008). One can easily envision the need to simultaneously optimize several objectives from intrinsic stability and specificity of folding to cost and yield of production to therapeutic safety and efficacy.

A number of strategies have emerged in the field for dealing with the inherent multi-objective nature of protein design problems. The simplest is to ignore alternate objectives and focus on one, such as stability. This is the nature of “forward design”, or positive design (Fleishman and Baker 2012), which seeks to optimize the stability of the target state. The key thermodynamic parameter that determines the probability a sequence will fold into the target state is the energy gap, i.e., the difference in stability between the target state and competing states (Fig. 1). Optimizing this gap is referred to as “negative design”. In forward design, the negative design problem is not explicitly considered under the assumption that improving the target stability will simultaneously improve the energy gap. This is often a reasonable approximation, particularly when the starting point for design is a natural protein, where much of negative design has been performed by evolution. On the other hand, there are examples where the positive design approach is ineffective, as seen in model systems (Yue et al. 1995), and in practical designs, such as the case of the development of collagen mimetic peptides (CMPs) presented later.

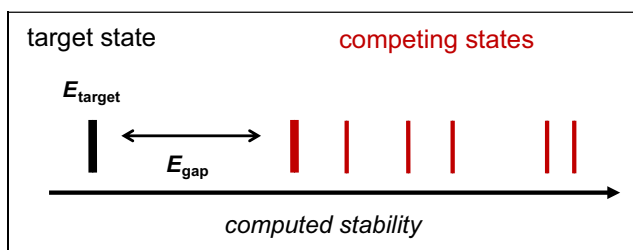


Fig. 1 Computed stabilities of a series of states representing target and competing protein conformations or binding interactions can be represented as an energy level diagram. In positive design, the stability of the target state E_{target} is the only objective that is optimized. For negative design, the difference in stability between target and competing states is also included to optimize the sequence to both stably and specifically fold in the target conformation

An alternative in multi-state design is to explicitly consider a relevant subset of conformational states that a sequence may adopt. This may be accomplished by combining states into a single objective, such as the Boltzmann probability of forming the target over competing states (Seeman and Kallenbach 1983; Havranek and Harbury 2003; Stapleton et al. 2015). In the case where multiple target states are desired, summing the stabilities of the sequence mapped onto each target structure (Ambroggio and Kuhlman 2006) or by separating sequence optimization for a set of states while gradually enforcing coupling constraints to drive convergence to a single solution (Sevy et al. 2015). Algorithmic solutions to explicitly treating an arbitrary number of competing states was used to develop an interactome of α -helical oligomers (Grigoryan et al. 2009).

Multi-objective problems may also take the form of disparate properties that are not scored using the same scales, units, or computational methods. One approach would be to combine multiple objectives into a static cost function that weights them appropriately. For example, in the de novo design of a four-helix bundle porphyrin cofactor binding protein, there were multiple design objectives: optimizing metal binding, protein sequence, and topology of the bundle. Metal coordination is highly sensitive to bond lengths and angles and is most effectively modeled using quantum mechanics calculations. To simplify this calculation, coordination geometry was expressed as a series of harmonic constraints on key bond lengths and angles, and concurrently optimized with sequence and topology using standard atomistic force fields (Cochran et al. 2005). The other approach is to algorithmically separate the quantum mechanics optimization of the active site configuration, choice of scaffold, and sequence optimization into discrete steps, as has been done in the development of several artificial enzymes (Zanghellini et al. 2006; Jiang et al. 2008; Röthlisberger et al. 2008). Fleishman and colleagues have explored using fuzzy logic operators as a strategy for incorporating disparate variables into a single scoring function (Warszawski et al. 2014). Importantly, this aggregation

approach has underlying mathematical limitations, such that it does not reach all available solutions in certain scenarios (Das and Dennis 1997).

Pareto optimality

Multi-objective tradeoff problems are not limited to protein engineering, and are found in nearly all fields, from economics to engineering to biological evolution. Unifying these problems is the concept of finding solutions that satisfy the Pareto optimality criterion. These are the subset of solutions where the evaluation of one objective could not be improved without reducing another; for example, the distribution of a fixed number of goods between two parties (Fig. 2). Formal definitions of the Pareto optimal set with regard to protein design can be found in previous studies (Belure et al. 2017). Algorithms that search for Pareto optimal solutions have been applied across science from materials (Hartke 2004), nanotechnology (Wiecha et al. 2017), and protein folding (Cutello et al. 2006) to understanding constraints on animal evolution (Sheftel et al. 2013).

The Pareto non-dominated set, often called the Pareto frontier, are attractive targets for laboratory characterization, particularly if the relative significance of individual objectives to design success is a priori unknown. Thus, several groups have developed algorithms for identifying the Pareto frontier in protein design problems. Bailey-Kellogg and colleagues developed PEPFR (Protein Engineering Pareto Frontier), which uses dynamic programming to implement an efficient divide-and-conquer approach, and applied it to design problems in therapeutic protein deimmunization, characterizing interacting sets of bZIP helical oligomers and optimizing site-directed recombination protocols for generating diverse libraries that maintain stability and activity (He et al. 2012; Salvat et al. 2015). Pareto refinement methods have also been

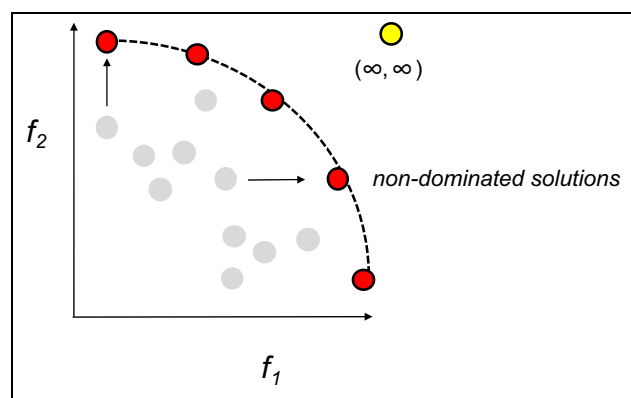


Fig. 2 The Pareto optimal set is depicted for a bi-objective problem f_1 , f_2 , where non-dominated solutions (red) cannot be further improved for f_1 without degrading f_2 or vice versa. Other solutions (gray) may be improved along either objective without compromising the other

applied to the design of stabilizing mutations to proteins that minimally disrupt the native structure, concurrently optimizing energy and RMSD from the initial structure (Nivón et al. 2013). State-of-the-art multi-objective optimization evolutionary algorithms such as SMS-EMOA have been applied to the design of peptide ligands that bind with reasonable affinity and selectivity for a specific isoform of 14-3-3 proteins (Sanchez-Faddeev et al. 2012). In characterizing protein designs that are well distributed along a Pareto optimal set, one is able to evaluate the relative importance of objectives to design success. Below, we discuss how the analysis of Pareto optimality may help guide experiments in the design of a collagen peptide interactome.

Specific case study: collagen peptide interactome

We have been using CMPs as model systems for exploring the tradeoff between stability, specificity, and system complexity in oligomeric interactions. Inside of cells, thousands of proteins co-exist and function at very high concentrations without non-specifically aggregating. The same phenomenon occurs outside of cells in the extracellular matrix, where complex mixtures of fibrous proteins, proteoglycans, and other biopolymers co-assemble into a structurally controlled network. We have been using computationally designed mimics of one such extracellular protein, collagen, to understand how specificity is maintained under such complex conditions.

Natural fibrous collagen is composed of three chains that associate into a triple-stranded helix. Type 1 collagen (COL1), a major component of skin and bone connective tissue, exists as a heterotrimer, containing two strands coming from the COL1A1 gene and one from the COL1A2 gene. This heterospecificity is largely governed by interactions between globular pro-domains which are cleaved during protein maturation. However, peptide fragments of COL1 show preference for heterotrimer formation in the absence of pro-domains (Saccà et al. 2002) and it is possible to generate heterotrimers composed of chains with three different sequences using networks of complementary charge pair interactions (Gaub and Hartgerink 2007; Fallas et al. 2009).

Subsequently, we demonstrated that combining stability and energy gap specificity as separate steps in a Monte Carlo simulated annealing (MCSA) protocol could produce an *abc* heterotrimer where assembly of a folded triple-helix required the presence of all three peptides (Fig. 3) (Xu et al. 2010, 2011). The target *abc* was maintained by an extensive network of interchain electrostatic interactions. Competing states such as *aaa*, *bbb*, and *ccc* homotrimers, or *aab*, *bbc* ... heterotrimers were disfavored by repulsive electrostatic interactions. Numerical simulations showed a clear tradeoff between heterotrimer stability and specificity. The most stable collagen peptide is rich in proline and hydroxyproline, cyclic

sidechain amino acids that provide conformational stability to the collagen triple-helix. However, these lack the charge pair interactions that promote specificity.

Each run of the MCSA on *abc* heterotrimers yielded a unique design with similar values for the stability and specificity objectives. This high degeneracy in objective space led us to consider a third objective: could multiple *abc*-type heterotrimers assemble when present in the same solution, i.e., could the complexity of the system be optimized within the design constraints? Mimicking the natural process of protein circular permutation, we generated additional peptides *e*, *f*, and *g*, which, when combined with *a*, *b*, and *c*, formed two separate heterotrimers: *abc + def* (Xu et al. 2013). However, the specificity was notably affected, suggesting an emerging tradeoff between specificity and complexity.

A systematic computational analysis of complexity, stability, and specificity demonstrates that complexity severely constrains the specificity of association (Fig. 4). The *abc* and *abc + def* designs lie on or near the apparent Pareto frontier, although the shallow dependence of energy gap and stability suggest that target stability could be increased without significant tradeoff for specificity. Increasing the number of peptides beyond six exhibits a significant tradeoff between complexity and specificity. Given that the number of states scales with the cube of the number of peptide types, the observed tradeoff is not surprising. Recently, we have applied a number of evolutionary and non-evolutionary algorithms to exploring the Pareto frontier for the 12-peptide system (Belure et al. 2017). Algorithms such as SMS-EMOA produce sets of non-dominated solutions, although replica exchange is also effective.

Conclusions

The collagen peptides provide a useful system for exploring tradeoffs among various objectives in designing an interactome of synthetic peptides. Notably, a similar constraint exists in natural proteomes, where complexity is limited by non-specific interactions between proteins (Tompa and Rose 2011), and measures of protein “stickiness” negatively correlate with expression levels (Levy et al. 2012). Nearly all protein design problems are either implicitly or explicitly multi-objective, and the development of efficient algorithms for producing a Pareto optimal set of sequences is an important goal. A practical approach for addressing multiple objectives in design would involve the synthesis and characterization of sequences that span a significant portion of the Pareto frontier in order to evaluate the relative importance of the multiple objectives. Subsequent design would then focus on libraries of solutions in a particular region of the frontier that represent an effective balance of the various objectives.

Compliance with ethical standards

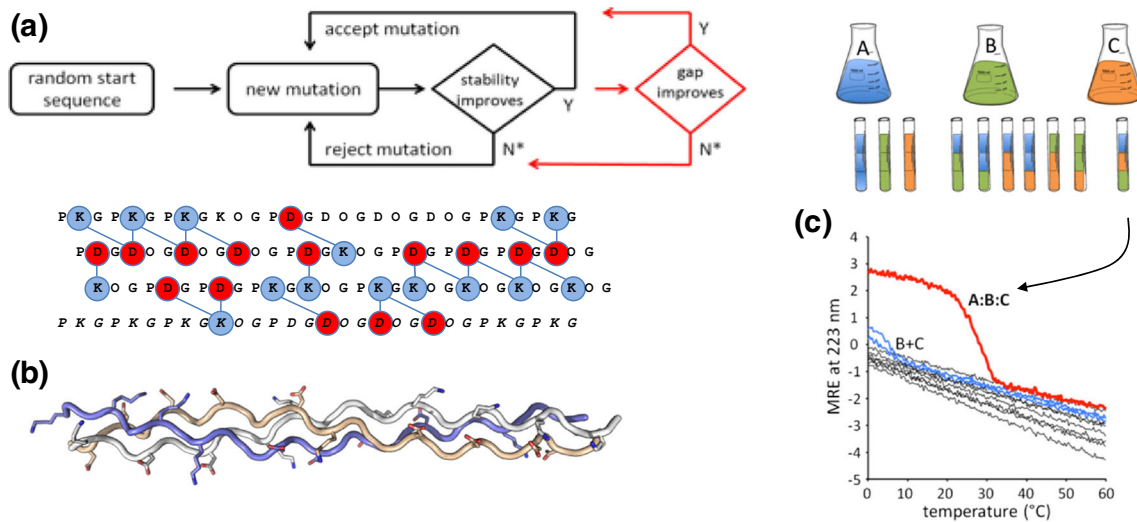


Fig. 3 Multi-objective design of an *abc* collagen heterotrimer. **a** Monte Carlo simulated annealing (MCSA) algorithm for collagen sequence design splitting the objective across two tiers: evaluating the stability of the target (*black*) and evaluating the gap between the target and competing states (*red*). **b** Top: example of designed sequences for peptides *a*, *b*, and *c* highlighting favorable charge pair interactions between acidic (*red*) and basic (*blue*) amino acids. Bottom: three-

dimensional model of that sequence on a collagen triple-helix. **c** Experimental evaluation of the design. Stock solutions of peptides *a*, *b*, and *c* are combined in various stoichiometric ratios. For a successful design, only *a* + *b* + *c* should form a triple-helix, where cooperative denaturation by circular dichroism spectroscopy (MRE: mean residue ellipticity) can be observed. Experimental details can be found in Xu et al. (2011)

Conflict of interest Vikas Nanda declares that he has no conflict of interest. Sandeep V. Belure declares that he has no conflict of interest. Ofer M. Shir declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

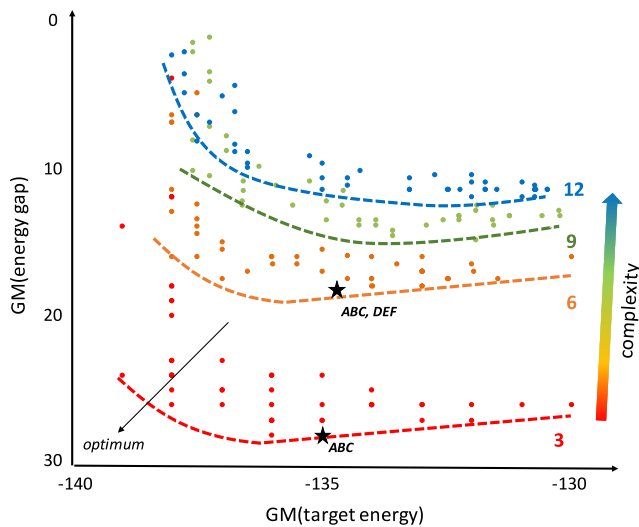


Fig. 4 Solution sets for multiple runs of a replica exchange algorithm (see Belure et al. 2017 for method details) for four complexity levels: *abc* (*red*), *abc* + *def* (*orange*), *abc* + *def* + *ghi* (*green*), and *abc* + *def* + *ghi* + *jkl* (*blue*) collagen peptide interactive designs. Visually estimated Pareto frontiers are indicated with *dotted lines*. Target energies and gaps represent the geometric mean (GM) of individual target heterotrimers at each level of complexity. Experimentally characterized solutions for *abc* (Xu et al. 2011) and *abc* + *def* (Xu et al. 2013) are indicated by *black stars*

References

Ambroggio XI, Kuhlman B (2006) Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J Am Chem Soc* 128(4):1154–1161

Belure SV, Shir OM, Nanda V (2017) Protein design by multiobjective optimization: evolutionary and non-evolutionary approaches. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2017), Berlin, Germany July 2017. ACM Press, New York

Braxton S, Wells JA (1992) Incorporation of a stabilizing calcium-binding loop into subtilisin BPN. *Biochemistry* 31(34):7796–7801

Camacho CJ, Katsumata Y, Ascherman DP (2008) Structural and thermodynamic approach to peptide immunogenicity. *PLoS Comput Biol* 4(11):e1000231

Chan HS, Dill KA (1991) “Sequence space soup” of proteins and copolymers. *J Chem Phys* 95(5):3775–3787

Choi Y, Hua C, Sentman CL, Ackerman ME, Bailey-Kellogg C (2015) Antibody humanization by structure-based computational protein design. *MAbs* 7(6):1045–1057

Cochran FV, Wu SP, Wang W, Nanda V, Saven JG, Therien MJ, DeGrado WF (2005) Computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. *J Am Chem Soc* 127(5):1346–1347

Cutello V, Narzisi G, Nicosia G (2006) A multi-objective evolutionary approach to the protein structure prediction problem. *J R Soc Interface* 3(6):139–151

Das I, Dennis JE (1997) A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Struct Multidiscip Optim* 14(1):63–69

Fallas JA, Gauba V, Hartgerink JD (2009) Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J Biol Chem* 284(39):26851–26859

Fleishman SJ, Baker D (2012) Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* 149(2):262–273

- Fromer M, Shifman JM (2009) Tradeoff between stability and multispecificity in the design of promiscuous proteins. *PLoS Comput Biol* 5(12):e1000627
- Gaub V, Hartgerink JD (2007) Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J Am Chem Soc* 129(9):2683–2690
- Grigoryan G, Reinke AW, Keating AE (2009) Design of protein-interaction specificity affords selective bZIP-binding peptides. *Nature* 458(7240):859–864
- Griswold KE, Bailey-Kellogg C (2016) Design and engineering of deimmunized biotherapeutics. *Curr Opin Struct Biol* 39:79–88
- Hall MP (2014) Biotransformation and in vivo stability of protein biotherapeutics: impact on candidate selection and pharmacokinetic profiling. *Drug Metab Dispos* 42(11):1873–1880
- Handel TM, Williams SA, DeGrado WF (1993) Metal ion-dependent modulation of the dynamics of a designed protein. *Science* 261(5123):879–885
- Hartke B (2004) Application of evolutionary algorithms to global cluster geometry optimization. *Appl Evol Comput Chem* 110:33–53
- Havranek JJ, Harbury PB (2003) Automated design of specificity in molecular recognition. *Nat Struct Mol Biol* 10(1):45–52
- He L, Friedman AM, Bailey-Kellogg C (2012) A divide-and-conquer approach to determine the Pareto frontier for optimization of protein engineering experiments. *Proteins Struct Funct Bioinf* 80(3):790–806
- Hellinga HW, Richards FM (1994) Optimal sequence selection in proteins of known structure by simulated evolution. *Proc Natl Acad Sci U S A* 91(13):5803–5807
- Howell SC, Inampudi KK, Bean DP, Wilson CJ (2014) Understanding thermal adaptation of enzymes through the multistate rational design and stability prediction of 100 adenylate kinases. *Structure* 22(2):218–229
- Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retro-aldol enzymes. *Science* 319(5868):1387–1391
- Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN (2013) Computational enzyme design. *Angew Chem Int Ed Engl* 52(22):5700–5725
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302(5649):1364–1368
- Levy ED, De S, Teichmann SA (2012) Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc Natl Acad Sci U S A* 109(50):20461–20466
- Nautiyal S, Woolfson DN, King DS, Alber T (1995) A designed heterotrimeric coiled coil. *Biochemistry* 34(37):11645–11651
- Nivón LG, Moretti R, Baker D (2013) A Pareto-optimal refinement method for protein design scaffolds. *PLoS One* 8(4):e59004
- Povolotskaya IS, Kondrashov FA (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* 465(7300):922–926
- Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195
- Saccà B, Renner C, Moroder L (2002) The chain register in heterotrimeric collagen peptides affects triple helix stability and folding kinetics. *J Mol Biol* 324(2):309–318
- Salvat RS, Parker AS, Choi Y, Bailey-Kellogg C, Griswold KE (2015) Mapping the Pareto optimal design space for a functionally deimmunized biotherapeutic candidate. *PLoS Comput Biol* 11(1):e1003988
- Sanchez-Faddeev H, Emmerich MTM, Verbeek FJ, Henry AH, Grimshaw S, Spink HP, van Vlijmen HW, Bender A (2012) Using multiobjective optimization and energy minimization to design an isoform-selective ligand of the 14-3-3 protein. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7610:12–24
- Seeman NC, Kallenbach NR (1983) Design of immobile nucleic acid junctions. *Biophys J* 44(2):201–209
- Sevy AM, Jacobs TM, Crowe JE Jr, Meiler J (2015) Design of protein multi-specificity using an independent sequence search reduces the barrier to low energy sequences. *PLoS Comput Biol* 11(7):e1004300
- Sheftel H, Shoval O, Mayo A, Alon U (2013) The geometry of the Pareto front in biological phenotype space. *Ecol Evol* 3(6):1471–1483
- Shifman JM, Mayo SL (2002) Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* 323(3):417–423
- Sormanni P, Aprile FA, Vendruscolo M (2015) The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol Biol* 427(2):478–490
- Stapleton JA, Whitehead TA, Nanda V (2015) Computational redesign of the lipid-facing surface of the outer membrane protein OmpA. *Proc Natl Acad Sci U S A* 112(31):9632–9637
- Summa CM, Rosenblatt MM, Hong JK, Lear JD, DeGrado WF (2002) Computational de novo design, and characterization of an A(2)B(2) diiron protein. *J Mol Biol* 321(5):923–938
- Tompa P, Rose GD (2011) The Levinthal paradox of the interactome. *Protein Sci* 20(12):2074–2079
- Voigt CA, Gordon DB, Mayo SL (2000) Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design. *J Mol Biol* 299(3):789–803
- Warszawski S, Netzer R, Tawfik DS, Fleishman SJ (2014) A “fuzzy”-logic language for encoding multiple physical traits in biomolecules. *J Mol Biol* 426(24):4125–4138
- Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 30(6):543–548
- Wiecha PR, Arbouet A, Girard C, Lecestre A, Larrieu G, Paillard V (2017) Evolutionary multi-objective optimisation of colour pixels based on dielectric nano-antennas. *Nat Nanotechnol* 12(2):163–169
- Xu F, Zhang L, Koder RL, Nanda V (2010) De novo self-assembling collagen heterotrimers using explicit positive and negative design. *Biochemistry* 49(11):2307–2316
- Xu F, Zahid S, Silva T, Nanda V (2011) Computational design of a collagen A:B:C-type heterotrimer. *J Am Chem Soc* 133(39):15260–15263
- Xu F, Silva T, Joshi M, Zahid S, Nanda V (2013) Circular permutation directs orthogonal assembly in complex collagen peptide mixtures. *J Biol Chem* 288(44):31616–31623
- Yue K, Fiebig KM, Thomas PD, Chan HS, Shakhnovich EI, Dill KA (1995) A test of lattice protein folding algorithms. *Proc Natl Acad Sci U S A* 92(1):325–329
- Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Röthlisberger D, Baker D (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15(12):2785–2794