Computational protein design with backbone plasticity

James T. MacDonald[1,2] and Paul S. Freemont[1,2]

[1]Centre for Synthetic Biology and Innovation, South Kensington Campus, London SW7 2AZ, U.K. and [2]Section of Structural Biology, Department of Medicine, Imperial College London, London SW7 2AZ, U.K.

Correspondence: James T. MacDonald (j.macdonald@imperial.ac.uk) or Paul S. Freemont (p.freemont@imperial.ac.uk)

The computational algorithms used in the design of artificial proteins have become increasingly sophisticated in recent years, producing a series of remarkable successes. The most dramatic of these is the *de novo* design of artificial enzymes. The majority of these designs have reused naturally occurring protein structures as 'scaffolds' onto which novel functionality can be grafted without having to redesign the backbone structure. The incorporation of backbone flexibility into protein design is a much more computationally challenging problem due to the greatly increased search space, but promises to remove the limitations of reusing natural protein scaffolds. In this review, we outline the principles of computational protein design methods and discuss recent efforts to consider backbone plasticity in the design process.

Introduction

A variety of different strategies have been developed to engineer novel globular proteins. These range from directed evolution, simple residue patterning methods, to atomic-level computational protein design. There has been less progress in the design of membrane proteins due to the difficulty in experimental characterisation [1], so this review concentrates mainly on the design of globular domains. Directed evolution methods are well established and have produced notable successes [2]. These methods generally require a starting protein sequence with some initial activity from which to generate and select variants. Mutations that increase the desired activity may be very rare, requiring high-throughput screening. The rational design of proteins using residue patterning has been particularly successful in the design of *de novo* helical bundle proteins [3], self-assembling coiled-coil peptides [4], and repeat proteins [5]. These proteins have been functionalised by intuitive manual design to introduce chemical activity [6,7].

In contrast with the previously described methods, computational protein design algorithms construct detailed full-atom models. The ability to place chemical moieties with atomic-level precision enables applications not possible with other protein engineering methods. Initial computational protein design work focussed on finding optimal sequences for fixed-backbone scaffolds taken from natural proteins [8–10]. These fixed-backbone computational design algorithms have been extended to introduce novel functionality, such as binding sites [11], libraries of fluorescent proteins [12], and *de novo* designed enzymes that catalyse reactions not found in nature [13,14]. In parallel to these developments, entirely *de novo* proteins, consisting of mainly canonical secondary structure and minimal loops, have been created by assembling backbone fragments from known protein structures, followed by iterated sequence design using the fixed-backbone approximation and energy minimisation [15–17]. However, in general, the rules, governing the designability of a given arbitrary backbone conformation, are not well understood. An outline of a typical computational protein design process is shown in Figure 1.

Given the rough landscape of full-atom potential energy functions, sequence design on a fixed-backbone artificially restricts the possible amino acid residues capable of being accommodated at a given position. Even small changes in backbone conformation may permit residues that were
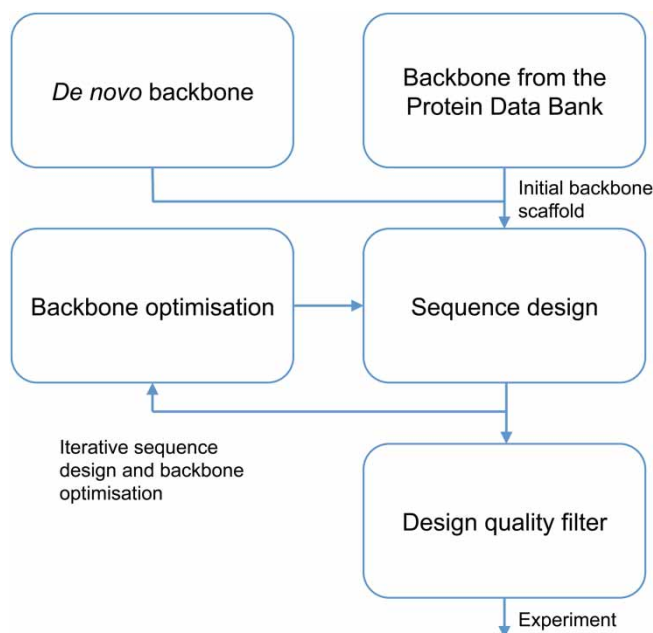
Figure 1. A typical computational protein design workflow.

Initial backbone structures can be either generated *de novo* or taken from solved protein structures. Sequences that stabilise the designed backbone structure are then computationally designed, and the backbone may be permitted to move as part of an iterative design cycle. Finally, promising designs are selected for experimental characterisation.

previously sterically hindered, and therefore improve the diversity in designed sequences [18]. The incorporation of larger scale backbone plasticity in computational designs will also allow further optimisation of engineered proteins [19], and the greater freedom will allow the implementation of more complex functionalities. For these reasons, flexible backbone protein design is an increasingly important area of research.

Computational protein design with the fixed-backbone approximation: the inverse folding problem

Computational protein design can be considered to be composed of two linked problems. The first problem is selecting or generating a plausible and designable backbone conformation. The second problem is finding sequences capable of specifically stabilising that backbone structure, also known as the inverse folding problem. The latter of these two problems will be discussed in this section.

Most modern methods use full-atom models of protein structure and molecular mechanics potential energy functions, consisting of a sum of covalent and non-covalent terms. These potential energy functions are often derived from force-fields developed for molecular dynamics simulations and may be supplemented with additional statistically derived terms [20]. Given a fixed-backbone, only side-chain identities and conformations are allowed to vary. Discrete libraries of side-chain conformations, known as rotamers [21], are commonly used to simplify the sampling and computation of potential energies. While the potential energy functions used in protein design are far from perfect, it has been observed that structures close to the native state almost always have the lowest potential energies, and it appears that conformational sampling is the bigger problem [22,23].

Stable proteins have a large energy gap between their native structure and all other possible structures. To rigorously determine whether a particular sequence specifically stabilises a given conformational state, it is necessary to evaluate the potential energy of that sequence over all possible (backbone and side-chain) conformations to calculate the partition function (i.e. the normalising constant required to ensure a probability distribution sums to 1). It is, then, possible to calculate the probability of any particular conformational state being occupied using the Boltzmann distribution (eqn 1, where P_i is the probability of the system being in state i, E_i is the energy of state i, k is the Boltzmann constant, T is the temperature, and the denominator is the partition

function, a sum over all states). Optimising this probability by searching sequence space would then solve the protein design problem for a given backbone conformation. Unfortunately, this is computationally intractable, so, in practice, a variety of approximations have been used. Most common methods approach this problem by optimising a potential energy function by trialling different side-chain identities and rotamers without explicitly considering alternative backbone conformational states. It has been proposed that it is less important to consider alternative conformations in three dimensions as most low-energy decoys will have dissimilar structures, making mutations that stabilise the native state unlikely to stabilise the other conformations [24]. Side-chain identity and rotamer search may be conducted by deterministic methods, such as dead-end elimination [25,26], or by stochastic methods [27–29].

$$P_i = \frac{e^{(-E_i)/kT}}{\sum_{j=1}^{N} e^{(-E_j)/kT}} \tag{1}$$

Previously, it was found that proteins designed using hydrophobic patterning methods, did not seem to fold into well-defined native states and appeared to be more similar to molten globules. For this reason, early computational work concentrated on improving the specific packing in the hydrophobic cores of proteins [8]. This was followed by the ground-breaking automated sequence redesign of an entire small protein [9]. A larger scale test showed that this approach could successfully produce well-folded proteins by redesigning a range of different proteins [10]; however, it was notable that the redesigned proteins composed primarily of β-sheets appeared to be aggregated or unfolded.

The fixed-backbone assumption has proved to be sufficient to successfully create proteins with novel functionality without considering backbone flexibility. If a constellation of side-chain chemical groups can be defined that are predicted to carry out a given function (e.g. a transition state model), it is then possible to search existing protein structures for backbone positions capable of hosting this geometric arrangement, while taking into account side-chain degrees of freedom and steric clashes. Various algorithms have been developed to accomplish this [30,31]. The RosettaMatch algorithm uses an 'outside-in' approach by constructing the transition state model at the ends of each catalytic side-chain rotamer at all possible positions in the scaffold and recording the six-dimensional position of the transition state model in a hash table. If all catalytic geometric constraints can be satisfied simultaneously with a given selection of residue positions, this would result in the transition state model being reconstructed in the same position from all catalytic side-chain residues. Hits can be rapidly determined by scanning the hash table [31]. This algorithm enables the search of a large database of potential scaffolds, and its utility was dramatically demonstrated by the successful design of *de novo* enzymes using theoretic transition state models [13,14].

Backbone sampling methods in protein design

Despite the achievements of fixed-backbone design, it is clear that this approximation is not sufficient to accurately sample sequence space and, more importantly, greatly limits the opportunities to optimise functional interactions. Backbone motion is also known to be functionally important in many natural proteins in molecular recognition [32,33] and enzymes [34].

The active site search algorithms described in the previous section are only able to search putative scaffolds for three to four catalytic residue geometries which are probably not enough to recapitulate extraordinary catalytic activities of natural enzymes [35]. The ability to redesign backbone structures around the catalytic site is likely to offer opportunities to optimise enzymes in ways that are not available to fixed-backbone approaches. In a previous paper, Foldit players were able to redesign a 24-residue backbone section and increase the activity of an artificially designed Diels–Alderase enzyme >18-fold [19].

The consideration of backbone plasticity in protein design requires the sampling of both backbone conformational space and side-chain identities/rotamers, and this enormously expands the search space. Additionally, unlike side-chains, backbone conformations are not amenable to discretisation. For these reasons, initial work on flexible backbone design was based on parameterised coiled-coil backbones [36], rigid body movements of secondary structural elements [37], and the introduction of small random backbone dihedral angle perturbations during the design process [38].

In parallel to the advances in computational protein design, a number of groups working in the related field of protein structure prediction found that short backbone fragments taken from previously solved protein

structures could be used to explore backbone conformational space in an efficient way [39]. The backbone fragments are defined in terms of internal dihedral co-ordinates then as part of a Monte Carlo search procedure, random sections of the backbone are replaced with dihedral angles from the fragment in process called 'fragment insertion.'

In an extraordinary achievement, this fragment insertion process was used to assemble an entirely *de novo* backbone fold not observed in nature with a computationally design sequence. A high-resolution crystal structure confirmed that the protein did indeed fold into the designed structure with atom-level precision [15]. This approach was subsequently generalised and extended to other folds using emergent rules [16,17]. These *de novo* folds consisted of idealised secondary structural elements linked with loops of minimal length. The use of existing backbone fragments in this way ensures that the local structural features of the designed protein replicate those observed in real proteins and increases the chance that the new backbone structures are designable. Recently, these computational design methods have been successfully applied to the design of more complex artificial coiled-coil proteins [40–43].

Fragment insertion is a non-local move as replacing dihedral angles in a particular backbone segment results in a move that propagates down the entire polypeptide chain. This is an inherently highly disruptive move, resulting in low acceptance rates in Monte Carlo simulations. However, fragment insertion can be turned into a local move by combining it with methods that can close chain breaks. A fragment is inserted midway along the chain, and a break is introduced at the N- or C-terminal end of the insertion so that the rest of the polypeptide chain is not moved. Adjustments then need to be made to the dihedral angles in the fragment such that the chain recloses. There are various algorithms that solve this loop closure problem, many of which are related to methods used to control robotic arms. These include cyclic co-ordinate descent (CCD), where each backbone dihedral angle is optimised in turn until the correct geometry at the break is restored [44], kinematic closure, where all dihedral angles in the loop may be freely varied except six dihedrals which are solved for loop closure using polynomial resultants [45], and stochastic closure methods [46].

Fragment-based approaches have been used to computationally design loop structures on natural protein scaffolds. By selecting backbone fragments from the PDB with endpoints that superimpose with the anchor residues in the scaffold, Hu et al. were able to graft 10-residue loops on to the protein tenascin. The inserted loops ranged from 0.9 to 1.6 Å backbone RMSD from the wild-type loop. The loop endpoints were close enough to the anchor residues that the loops could be closed by gradient minimisation. Two loops were solved using X-ray crystallography, and one was found to match the designed loop conformation with sub-Angstrom RMSD [47]. CCD has been used together with fragment insertion to design a *de novo* loop that alters the substrate specificity of an enzyme [48]. In this work, short backbone fragments were inserted before and after a fixed anchor residue predicted to alter substrate binding followed by CCD to close the chain breaks. This approach produced a design with a four-residue sequence change which was confirmed to be in the correct conformation by X-ray crystallography.

While larger scale backbone motions can be modelled using fragment insertion, more subtle backbone movements are also very important in protein modelling. Natural proteins can be quite tolerant to mutations as the backbone can adjust to accommodate side-chains that would not be permitted using the fixed-backbone approximation. Several methods have been developed to model small backbone perturbations. In many protein design applications, cycles of sequence design followed by potential energy minimisation of the whole structure (including the backbone) are carried out to permit some degree of backbone flexibility. Other methods include extensions to the dead-end elimination algorithm to include backbone flexibility [49]. While these methods result in designs with lower potential energies, they do not always recapitulate the natural sequence variation observed in these proteins. A novel local backbone move called 'backrub' was developed after inspecting very high-resolution crystal structures for alternative backbone conformations [50]. This move rotates the backbone around the axis connecting $C\alpha_{i-1}$ and $C\alpha_{i+1}$, followed by compensating rotations of the $C\alpha_{i-1}$ to $C\alpha_i$ and the $C\alpha_i$ to $C\alpha_{i+1}$ peptide bonds. This results in a shift in the direction of the central side-chain but with minimal changes to backbone hydrogen-bonding geometry. This method was generalised and implemented in the Rosetta software package and, when coupled to sequence design, was shown to significantly improve the recapitulation of experimentally observed sequence variation in protein–protein [51] and protein–ligand [52] interfaces compared with fixed-backbone design methods.

In the past few years, we have developed new algorithms and software for the fragment-free sampling of backbone loop conformations using a coarse-grained model [53]. This method uses a coarse-grained potential energy function [54] to rapidly sample plausible backbone conformations at the carbon-α level, then accurately
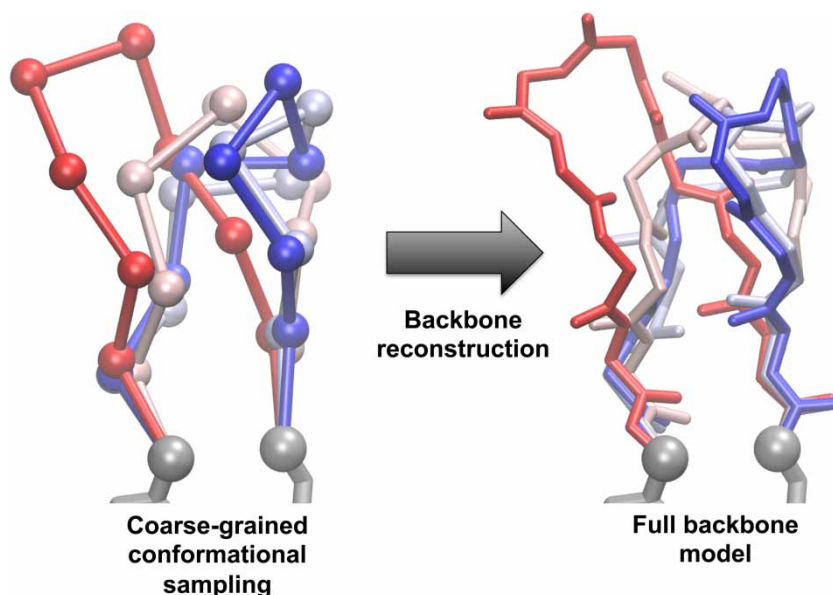
Figure 2. Sampling backbone loop conformations using a coarse-grained model.

Conformational space can be rapidly sampled using a reduced representation before being rebuilt into a full-atom model as part of a hierarchical design strategy. The grey atoms are the fixed anchor atoms at the N- and C-terminal ends of the loop being sampled, and the red/blue chains are alternative loop structures.

reconstructs the full backbone model using a structural alphabet derived using Gaussian mixture models [55]. The potential energy function consists of a pseudo hydrogen-bonding term, a soft steric repulsive term, a pseudo Cα-Cα bond term, and local structural terms. The local terms were derived using a structural alphabet and include a pseudo Cα-Cα-Cα bond angle term, a pseudo Cα-Cα-Cα-Cα dihedral angle term, and reference terms to ensure that the equilibrium distributions of each structural alphabet letter reproduce those observed in a high-resolution training set. An ensemble of loop conformations can be sampled by running successive simulated annealing Monte Carlo trajectories using only local moves that do not propagate down the rest of the chain (Figure 2). When side-chains were added, and the structure was energy gradient minimised using the Rosetta software package, this approach produced results that were equivalent to fragment insertion methods [53]. We propose that the ability to sample directly using coarse-grained potential energy functions enables the efficient incorporation of functional geometric restraints and the use of more sophisticated sampling methods that are more difficult to achieve with fragment insertion methods. Recently, we have successfully applied this fragment-free method to the design of *de novo* backbone protein design [56].

Conclusion

In the present paper, we have described the methodological advances in computational protein design in the past few decades. Initial approaches to computational design considered only fixed-backbone structures. It has become clear that the incorporation of backbone flexibility is essential to fully explore sequence space and to enable more complex designs. This backbone flexibility can range from small-scale motions, that permit slightly different side-chain orientations, to the large-scale redesign of complete sections of the protein backbone. The rules governing whether a given arbitrary backbone conformation is designable are not well understood. In particular, the computational design of *de novo* backbone loops has proved to be particularly challenging. There has been more success in the design of *de novo* folds composed of secondary structural elements and minimal loops as there are well-understood rules governing the packing of these elements. To date, the vast majority of successful computationally designed functional proteins have relied on fixed-backbone design methods. A better understanding of backbone designability and new design algorithms will enable the complete remodelling of large sections of the protein scaffold resulting in improved enzymes.

Abbreviation

CCD, cyclic co-ordinate descent.

Author Contribution

Both the authors contributed to writing this manuscript.

Funding

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1   Barth, P. and Senes, A. (2016) Toward high-resolution computational design of the structure and function of helical membrane proteins. *Nat. Struct. Mol. Biol.* **23**, 475–480 doi:10.1038/nsmb.3231

2   Romero, P.A. and Arnold, F.H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 doi:10.1038/nrm2805

3   DeGrado, W., Wasserman, Z. and Lear, J. (1989) Protein design, a minimalist approach. *Science* **243**, 622–628 doi:10.1126/science.2464850

4   Woolfson, D.N. (2005) The design of coiled-coil structures and assemblies. In *Advances in Protein Chemistry*, pp. 79–112

5   Javadi, Y. and Itzhaki, L.S. (2013) Tandem-repeat proteins: regularity plus modularity equals design-ability. *Curr. Opin. Struct. Biol.* **23**, 622–631 doi:10.1016/j.sbi.2013.06.011

6   Kaplan, J. and DeGrado, W.F. (2004) De novo design of catalytic proteins. *Proc. Natl Acad. Sci. USA* **101**, 11566–11570 doi:10.1073/pnas.0404387101

7   Koder, R.L., Anderson, J.L.R., Solomon, L.A., Reddy, K.S., Moser, C.C. and Dutton, P.L. (2009) Design and engineering of an O_2 transport protein. *Nature* **458**, 305–309 doi:10.1038/nature07841

8   Desjarlais, J.R. and Handel, T.M. (1995) De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018 doi:10.1002/pro.5560041006

9   Dahiyat, B.I. (1997) De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 doi:10.1126/science.278.5335.82

10 Dantas, G., Kuhlman, B., Callender, D., Wong, M. and Baker, D. (2003) A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **332**, 449–460 doi:10.1016/S0022-2836(03)00888-X

11 Tinberg, C.E., Khare, S.D., Dou, J., Doyle, L., Nelson, J.W., Schena, A. et al. (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 doi:10.1038/nature12443

12 Treynor, T.P., Vizcarra, C.L., Nedelcu, D. and Mayo, S.L. (2007) Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc. Natl Acad. Sci. USA* **104**, 48–53 doi:10.1073/pnas.0609647103

13 Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J. et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 doi:10.1038/nature06879

14 Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Rothlisberger, D., Zanghellini, A. et al. (2008) De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 doi:10.1126/science.1152692

15 Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L. and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 doi:10.1126/science.1089427

16 Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T. et al. (2012) Principles for designing ideal protein structures. *Nature* **491**, 222–227 doi:10.1038/nature11600

17 Lin, Y.-R., Koga, N., Tatsumi-Koga, R., Liu, G., Clouser, A.F., Montelione, G.T. et al. (2015) Control over overall shape and size in de novo designed proteins. *Proc. Natl Acad. Sci. USA* **112**, E5478–E5485 doi:10.1073/pnas.1509508112

18 Smith, C.A. and Kortemme, T. (2008) Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol.* **380**, 742–756 doi:10.1016/j.jmb.2008.05.023

19 Eiben, C.B., Siegel, J.B., Bale, J.B., Cooper, S., Khatib, F., Shen, B.W. et al. (2012) Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat. Biotechnol.* **30**, 190–192 doi:10.1038/nbt.2109

20 Shen, M.-y. and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 doi:10.1110/ps.062416606

21 Ponder, J.W. and Richards, F.M. (1987) Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791 doi:10.1016/0022-2836(87)90358-5

22 Bradley, P., Misura, K.M.S. and Baker, D. (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868–1871 doi:10.1126/science.1113801

23 Conway, P., Tyka, M.D., DiMaio, F., Konerding, D.E. and Baker, D. (2014) Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* **23**, 47–55 doi:10.1002/pro.2389

24 Shakhnovich, E.I. (1998) Protein design: a perspective from simple tractable models. *Folding Des.* **3**, R45–R58 doi:10.1016/S1359-0278(98)00021-2

25 Desmet, J., De Maeyer, M., Hazes, B. and Lasters, I. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539–542 doi:10.1038/356539a0

26 Dahiyat, B.I. and Mayo, S.L. (1996) Protein design automation. *Protein Sci.* **5**, 895–903 doi:10.1002/pro.5560050511

27   Lee, C. and Levitt, M. (1991) Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature* **352**, 448–451 doi:10.1038/352448a0

28   Hellinga, H.W. and Richards, F.M. (1994) Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl Acad. Sci. USA* **91**, 5803–5807 doi:10.1073/pnas.91.13.5803

29   Kuhlman, B. and Baker, D. (2000) Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA* **97**, 10383–10388 doi:10.1073/pnas.97.19.10383

30   Hellinga, H.W., Caradonna, J.P. and Richards, F.M. (1991) Construction of new ligand binding sites in proteins of known structure. II. Grafting of a buried transition metal binding site into *Escherichia coli* thioredoxin. *J. Mol. Biol.* **222**, 787–803 doi:10.1016/0022-2836(91)90511-4

31   Zanghellini, A., Jiang, L., Wollacott, A.M., Cheng, G., Meiler, J., Althoff, E.A. et al. (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 doi:10.1110/ps.062353106

32   James, L.C., Roversi, P. and Tawfik, D.S. (2003) Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367 doi:10.1126/science.1079731

33   Lange, O.F., Lakomek, N.-A., Fares, C., Schroder, G.F., Walter, K.F.A., Becker, S. et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 doi:10.1126/science.1157092

34   Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A. et al. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 doi:10.1038/nature04105

35   Baker, D. (2010) An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817–1819 doi:10.1002/pro.481

36   Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T. and Kim, P.S. (1998) High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 doi:10.1126/science.282.5393.1462

37   Su, A. and Mayo, S.L. (1997) Coupling backbone flexibility and amino acid sequence selection in protein design. *Protein Sci.* **6**, 1701–1707 doi:10.1002/pro.5560060810

38   Desjarlais, J.R. and Handel, T.M. (1999) Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.* **290**, 305–318 doi:10.1006/jmbi.1999.2866

39   Bowie, J.U. and Eisenberg, D. (1994) An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function. *Proc. Natl Acad. Sci. USA* **91**, 4436–4440 doi:10.1073/pnas.91.10.4436

40   Huang, P.-S., Oberdorfer, G., Xu, C., Pei, X.Y., Nannenga, B.L., Rogers, J.M. et al. (2014) High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 doi:10.1126/science.1257481

41   Thomson, A.R., Wood, C.W., Burton, A.J., Bartlett, G.J., Sessions, R.B., Brady, R.L. et al. (2014) Computational design of water-soluble α-helical barrels. *Science* **346**, 485–488 doi:10.1126/science.1257452

42   Brunette, T.J., Parmeggiani, F., Huang, P.-S., Bhabha, G., Ekiert, D.C., Tsutakawa, S.E. et al. (2015) Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 doi:10.1038/nature16162

43   Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B.L. et al. (2015) Rational design of α-helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 doi:10.1038/nature16191

44   Canutescu, A.A. and Dunbrack, R.L. (2003) Cyclic coordinate descent: a robotics algorithm for protein loop closure. *Protein Sci.* **12**, 963–972 doi:10.1110/ps.0242703

45   Coutsias, E.A., Seok, C., Wester, M.J. and Dill, K.A. (2006) Resultants and loop closure. *Int. J. Quantum Chem.* **106**, 176–189 doi:10.1002/qua.20751

46   Minary, P. and Levitt, M. (2010) Conformational optimization with natural degrees of freedom: a novel stochastic chain closure algorithm. *J. Comput. Biol.* **17**, 993–1010 doi:10.1089/cmb.2010.0016

47   Hu, X., Wang, H., Ke, H. and Kuhlman, B. (2007) High-resolution design of a protein loop. *Proc. Natl Acad. Sci. USA* **104**, 17668–17673 doi:10.1073/pnas.0707977104

48   Murphy, P.M., Bolduc, J.M., Gallaher, J.L., Stoddard, B.L. and Baker, D. (2009) Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl Acad. Sci. USA* **106**, 9215–9220 doi:10.1073/pnas.0811070106

49   Georgiev, I. and Donald, B.R. (2007) Dead-end elimination with backbone flexibility. *Bioinformatics* **23**, I185–I194 doi:10.1093/bioinformatics/btm197

50   Davis, I.W., Arendall, W.B., Richardson, D.C. and Richardson, J.S. (2006) The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure* **14**, 265–274 doi:10.1016/j.str.2005.10.007

51   Humphris, E.L. and Kortemme, T. (2008) Prediction of protein-protein interface sequence diversity using flexible backbone computational protein design. *Structure* **16**, 1777–1788 doi:10.1016/j.str.2008.09.012

52   Ollikainen, N., de Jong, R.M. and Kortemme, T. (2015) Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity. *PLoS Comput. Biol.* **11**, e1004335 doi:10.1371/journal.pcbi.1004335

53   MacDonald, J.T., Kelley, L.A. and Freemont, P.S. (2013) Validating a coarse-grained potential energy function through protein loop modelling. *PLoS ONE* **8**, e65770 doi:10.1371/journal.pone.0065770

54   MacDonald, J.T., Maksimiak, K., Sadowski, M.I. and Taylor, W.R. (2010) De novo backbone scaffolds for protein design. *Proteins: Struct., Funct., Bioinf.* **78**, 1311–1325 doi:10.1002/prot.22651

55   Moore, B.L., Kelley, L.A., Barber, J., Murray, J.W. and MacDonald, J.T. (2013) High-quality protein backbone reconstruction from alpha carbons using Gaussian mixture models. *J. Comput. Chem.* **34**, 1881–1889 doi:10.1002/jcc.23330

56   MacDonald, J.T., Kabasakal, B.V., Godding, D., Kraatz, S., Henderson, L., Barber, J. et al. (2016) Synthetic beta-solenoid proteins with the fragment-free computational design of a beta-hairpin extension. *Proc. Natl Acad. Sci. USA,* doi:10.1073/pnas.1525308113