Protein Design

Ragothaman M Yennamalli, Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India

Introduction

A cursory search of the Protein Data Bank (PDB) with the keyword "*de novo*" returns 962 entries, while for the keyword "designed", 6225 entries were returned [as of August 2017]. Although these numbers (*de novo* or designed structures) are 1000 times lesser in magnitude (compared to the total number of structures the PDB holds, which is close to 135,000 structures), these ~7000 structures indicate how far the field of protein design has advanced, since 1950s. Protein designing projects are ambitious in their goal due to the simple yet complex problem of protein folding. Much has been learned about how a protein folds, and these fundamental knowledge have helped further the area of protein design to conceive proteins with imaginative structures (Richardson and Richardson, 1989; Bowie *et al.*, 1991).

The impetus for protein design is two-fold: (i) assumption that we can design a complex natural system from first principles, and (ii) the "made to order" macromolecules that can solve important biochemical hurdles. The basic or fundamental problem with protein design in achieving a three dimensional, stable, and functional macromolecule is to cross the conformational entropy from the primary to the tertiary structure (Bowie *et al.*, 1991; Dahiyat and Mayo, 1997). There are many methods that can be used to reduce the conformational entropy (Baxa *et al.*, 2014). These include covalent cross-links and other artificial constraints that limit the conformational possibilities of the designed molecule (Leitner *et al.*, 2010; Sinz *et al.*, 2015).

There have been two basic design principles employed in designing proteins *de novo*: positive design and negative design (DeGrado *et al.*, 1989). Positive design of protein structures is the idea to design a protein with the desired structure as the goal, and rationally add/remove residues to achieve that structure. In contrast, negative design involves designing a structure along with ways to reduce formation of or competition from alternative conformations that may arise. While both methods involve rational design, there are advantages of using one over the other. Nevertheless, *de novo* design of protein structures using both methods has been successful to a varying degree. **Fig. 1** shows a representative set of structures that have been successfully designed and have advanced the field of protein design.

Since 1950s, designing alpha helices took precedence than the beta sheets, due to:

a) The stabilizing hydrogen bond network (**Fig. 2**).
b) The observation of isolated helices being stable in solution (Brown and Klee, 1971; Kim and Baldwin, 1984; Marqusee and Baldwin, 1987; Shoemaker *et al.*, 1987; Marqusee *et al.*, 1989).
c) Its oligomerization property observed by Crick, Pauling and Corey in supercoiled helices that when two helices twist around each other there are 3.5 residues per turn, which is less than the ideal 3.6 residues per turn rule, thereby leading to a repetition of the entire structure at every seven residues (Crick, 1953; Pauling and Corey, 1953).
d) The possibility of designing the minimal sequence, which can be repeated to construct a four- or six-bundle helices that can either self assemble or assemble in the presence of an external assembly inducer (Schafmeister *et al.*, 1997).

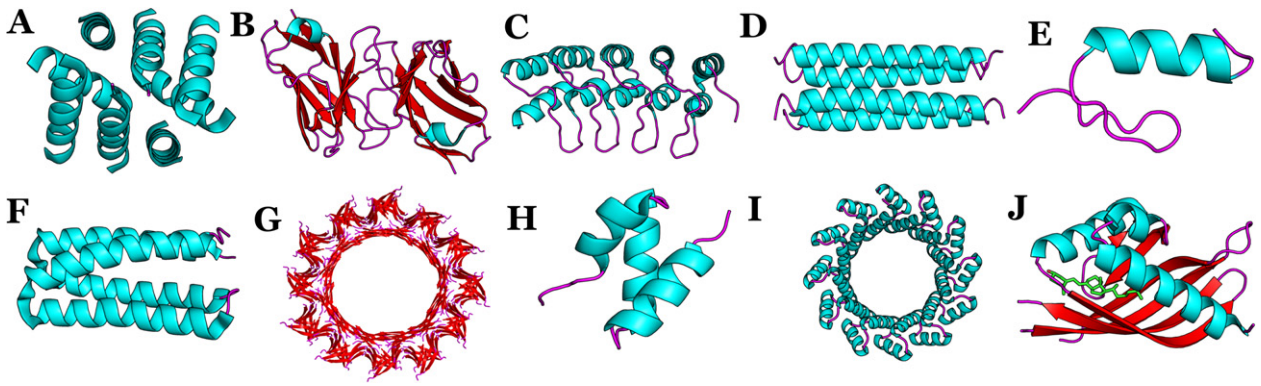

Fig. 1 A representative set of structures that have been designed *de novo*, where they are shown in cartoon representation with helices colored cyan, β-strands colored red, and loops colored magenta. (A) Octameric *de novo* Designed Peptide (pdb id:1l4x), (B) *de novo* Design of an Antibody Combining Site (pdb id:1ivl), (C) Self-Assembling Cyclic Protein Homo-Oligomer (pdb id:4hb5), (D) Right-Handed Coiled Coil Tetramer (pdb id:1rh4), (E) Beta Beta Alpha Protein Motif (pdb id:1fsv), (F) *de novo* Design of a Hyperstable Non-Natural Protein-Ligand Complex (pdb id:5tgw), (G) Giant Double-Walled Peptide Nanotube (pdb id:5vf1), (H) *de novo* Designed Mini Protein Hhh_Rd1_0142 (pdb id:5uoi), (I) Computationally Designed Left-Handed Alpha/Alpha Toroid With 12 Repeats (pdb id:5byo), and (J) Computationally Designed Vitamin-D3 Binder (pdb id:5iep).

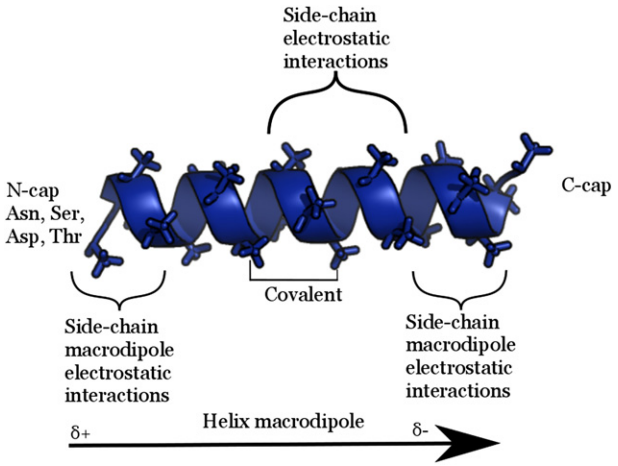

Fig. 2 Helix stabilizing properties. Schematic representation of an ideal helix indicating the various properties that stabilize the helix. Adapted from Bryson, J.W., *et al.*, 1995. Protein design: A hierarchic approach. Science (New York, NY) 270 (5238), 935–941.

Fundamentals

Before taking a sequence and designing it into a novel structure (α-helical or β-sheets or mixed), few pointers need to be kept in mind:

1. Similar stretch of residues can form different secondary structure.
   It has been observed that the same residue stretch of up to five amino acids can form different conformations, such as α-helix, β-strand, and loop. This indicates that a sequence forming a structure is dependent on its local environment.
2. There are only so many folds present.
   The relatively limited number of folds in the protein universe can be considered as a double edged sword. The 1000 odd folds we know today indicate that proteins tend to fold in one of the restricted ways. It also suggests the possibility of creating new folds using computational tools. The former has led to the reverse folding problem, called threading, where a sequence is checked for its compatibility to fold into one of the known folds.
3. Reducing the entropy of protein by introducing disulfide bonds (Wetzel, 1987).
4. Glycine introduces relatively more conformational freedom than the other 19 amino acids.
5. In contrast, Proline introduces less conformational freedom due to its covalent bond with the main chain.
6. It is known that the residues at the end of a helix are positively (at the C-terminus) and negatively (at the N-terminus) charged, which help in stabilizing the helix by balancing the helix dipole charge. Thus, changing the helix stabilizing residues influences the stability of a protein (Nicholson *et al.*, 1988; Richardson and Richardson, 1988).
7. If a core of a protein has a cavity, filling the cavity makes the protein destabilize it.

Designing an α-Helical Structure

Historically, alpha helical peptides and all-α proteins were relatively easier to design. In order to design alpha helical proteins, the sequence under consideration should be scrutinized with the following guidelines.

1. The sequence should be able to form amphiphilic secondary structures, where there is a periodicity of polar and nonpolar residues. For example, every third or fourth position being nonpolar.
2. A general rule of thumb is that the sequence should have residues that are highly likely to be part of a helix (Ala, Glu, Leu, and Met) (Richardson, 1981). Specifically, multiple alanine residues stabilize the helix. However, if the nonpolar periodicity of 3 or 4 is not maintained, then there is a lower preference to form a helical structure (Xiong *et al.*, 1995).
3. Introduction of salt-bridges and hydrogen bonds between side chains of residues that are one helical turn away (Marqusee *et al.*, 1989; Lyu *et al.*, 1990; Huyghues-Despointes *et al.*, 1993; Park *et al.*, 1993; Scholtz *et al.*, 1993).
4. A charged residue can be introduced at the N or C terminus of the α-helix, creating a macrodipole (**Fig. 2**).
5. Capping the helix ends with Asn, Ser, Asp, or Thr to satisfy the hydrogen-bond donors and acceptors.
6. Adding hydrogen bond(s) between side chains of residues that are one helical turn away.
7. In the case of Ser, Thr, and other amino acids that can be phosphorylated, their location in the interior of the helix may lead to destabilization. However, the N-terminus capping of a phosphorylated residue leads to stabilization of the helix. This is due to the electronic interaction of the phosphorylated residue with the peptide backbone.

Certain tools that can be used specifically to design helices are helical wheel and predicting helicity.

Helical wheel is a simple tool to identify if the distribution of charged residues and hydrophobic residues would lead to aggregation. When the helix has majorly hydrophobic or nonpolar residues on one side, there is a high chance of protein aggregation. This highly efficient tool has been used for designing proteins that are primarily coiled coil, and for helix bundles comprised of 3, 4, 5, and 6 helices. Specifically, the tool becomes effective when these helices are linked by loops or turns and to check if the "ridges in groove" or "knobs in holes" packing needs to be checked.

Some of the tools that plot a helical wheel for a given sequence are: DrawCoil 1.0 (see Relevant Websites Section), Pepwheel (see Relevant Websites Section), and Helical Wheel Projection (see Relevant Websites Section).

Also, the designed sequence can be checked for its propensity to form helices by submitting the sequence to AGADIR (see Relevant Websites Section) that calculates helicity (Munoz and Serrano, 1994, 1995a,b, 1997; Lacroix *et al.*, 1998).

Designing a β-Sheet Structure

β-sheets can be parallel, antiparallel, and mixed in nature. Due to absence of planarity, parallel β-sheets are less stable than antiparallel sheets. This is due to the relative absence of inter-strand hydrogen bonds in parallel β-sheets, which provide for stronger interaction. **Fig. 1(B), (G),** and **(J)** highlight the successful design of a β-strand rich proteins.

In 1996, the design of a decapeptide adopting a β-hairpin structure involving a β-bulge and three other dodecapeptides adopting a type I′ β-turn made the protein design possible for β-rich proteins (de Alba *et al.*, 1996; Ramirez-Alvarado *et al.*, 1996; Stanger and Gellman, 1998). These reports fueled enthusiasm in four different groups to design a three stranded antiparallel β-sheet proteins (each differing in residue length) (Kortemme *et al.*, 1998; Schenck and Gellman, 1998; Sharman and Searle, 1998; de Alba *et al.*, 1999). Irrespective of the method they employed, some of the criteria that were included in all the above mentioned β-sheet proteins are:

1. Residues that have higher β-strand propensities were included in the sequence.
2. Inter-strand pairs were selected that have higher preference to make a stable bond.
3. Involving positively charged residues (2 to 5 residues, at least) so that aggregation of proteins is eliminated, and increases solubility.

Positive charge distribution is essential in designing β-sheet proteins, because when they are distributed on both sides of the sheet, aggregation of proteins is reduced to a large extent. Another method to design a β-sheet protein with a stable hydrophobic core involves adding a type I′ β-turn to the designed β-hairpin protein by using residues, such as Phe, Trp, Asn, and Gly, so that the side chains face the other strand's hydrophobic residues' side chains (Tyr and Val) (Griffiths-Jones and Searle, 2000).

Similar to α-helices, some of the guidelines to be used while designing β-sheet proteins are as follows:

1. The role of a β-turn is crucial, as it dictates the β-strand. This is a necessary but not a sufficient condition while designing β-sheet proteins. For example, the D form of Pro in the turn stabilizes the β-hairpin structure, compared to the L form.
2. Residues that have high propensity to form β-sheet should be used. For example, Val, Ile, Phe, Trp, Tyr, Leu, and Thr.
3. Gly should be avoided, as its incorporation leads to destabilization.
4. Including salt-bridge forming residues in inter-strands stabilize the β-sheet protein. For example, a salt-bridge formed between Glu and Lys residues. On the same note, salt-bridges at the ends of β-hairpin are more stabilizing.
5. Presence of interactions between hydrophobic residues, such as Ile-Trp, Ser-Thr, and Trp/Val-Tyr/Phe stabilize the protein as their contributions are larger.
6. Side chain-side chain interactions from diagonal directions between two strands also contribute towards stability. For example, Tyr-Lys, Phe/Trp-Lys/Arg interactions.
7. Presence of a right-handed β-sheet twist.
8. The hydrophobic cluster between the strands contributed by Trp, Val, Phe, and Tyr residues vastly stabilize the β-sheet proteins.
9. As a general rule of thumb, incorporating disulfide bonds stabilize the protein with β-hairpin structures.

Relatively speaking, there are fewer examples of β-sheet proteins designed *de novo*. However, α/β mixed structures have been designed with great success (Struthers *et al.*, 1996). For example, Top7 by David Baker group and a $\beta\beta\alpha$ motif structure that is similar to a zinc-finger, which has a β-hairpin structure (Kuhlman *et al.*, 2003).

Tools Currently Available for Protein *de novo* Design

Irrespective of topology of the intended protein to be designed, some tools are listed below that are routinely used for *de novo* protein design, and also keeping in mind the ease of use from user's perspective.

Rosetta and Rosetta Design

Among the tools that are currently available, Rosetta (Simons *et al.*, 1999), developed by David Baker group at University of Washington, has been the most popular and widely used (Jiang *et al.*, 2008; Siegel *et al.*, 2010; Damborsky and Brezovsky, 2014). The suite of software (see Relevant Websites Section) has sped the design of protein structures. Details of Rosetta's design algorithm and specifics are discussed in Simons *et al.* (1997, 1999), Raveh *et al.* (2010) and DiMaio *et al.*, (2011) and they have been reviewed elsewhere (Mandell and Kortemme, 2009; Der and Kuhlman, 2011). Specifically, Rosetta Design (see Relevant Websites Section) can be accessed via the command line interface of Rosetta or via webserver (Lyskov *et al.*, 2013).

Evodesign

Evodesign (see Relevant Websites Section) developed by Zhang group at University of Michigan is a web based server to design protein sequences using a scaffold as an input. The scaffold is searched against the known protein families and the resulting conformation takes into consideration local environmental factors, such as solvent accessibility, packing, and secondary structure (Mitra *et al.*, 2013).

Protein WISDOM

Protein Wisdom (see Relevant Websites Section) is a web server tool for designing proteins from sequence information. The design and validation involves two steps: using either a rigid or flexible scaffold (or template), a sequence is selected from a pool of candidate sequences; and validation is done by fitting the sequence into known folds to calculate the "fold specificity" (Smadbeck *et al.*, 2013).

OSPREY

Open Source Protein REdesign for You (OSPREY) (see Relevant Websites Section) like Rosetta is a suite of programs developed by Donald group at Duke University. While Rosetta is licensed free for academics, OSPREY is open-source and freely available to download and use. OSPREY uses protein flexibility to create low-energy corpus of structures to identify the globally optimal structure. From the user's perspective, OSPREY runs as a standalone tool and is not available as a web-server (Gainza *et al.*, 2013).

ISAMBARD

Intelligent System for Analysis, Model Building And Rational Design (ISAMBARD) (see Relevant Websites Section) is another open-source suite of software for designing proteins developed by Woflsoon group at University of Bristol. Keeping with the popularity of Python, ISAMBARD uses predefined python object based method to design protein structures. Those with a basic python skill will be able to use this modular and scalable software to design protein structures (Wood *et al.*, 2017).

FireProt

Rather than a general protein design tool, FireProt (see Relevant Websites Section) is a specific protein design tool for designing multiple-point mutant proteins that are likely to be thermostable (Musil *et al.*, 2017).

iRDP

Similar to FireProt, in-silico Rational Design of Proteins (iRDP) (see Relevant Websites Section) is a webserver that uses a four-step approach to rationally design proteins. Specifically, from the input it compares existing protein structures for structural stability factors, followed by mutational analysis, and their impact to local environmental changes, and identifying the optimal structure that would have a higher thermostability than the previous structure (Panigrahi *et al.*, 2015).

IPRO

Iterative Protein Redesign and Optimization procedure (IPRO) (see Relevant Websites Section) from the Maranas group at Pennsylvania State University, uses a combinatorial approach to redesign a protein library using energy based scoring functions. As claimed by the developers, it uses a iterative process to make additive mutations that improve the designed protein's substrate specificity (Saraf *et al.*, 2006; Fazelinia *et al.*, 2007).

Scaffold Selection

Keeping in view that new protein can be designed from pre-existing structures; ScaffoldSelection (see Relevant Websites Section) is a tool that scans large sets of structures for a particular reaction scaffold (Malisi *et al.*, 2009). For the user, the tool can be downloaded as binaries for Linux, Mac, and Windows operating systems and run as a standalone software/tool.

Pocketoptimizer

Pocketoptimizer (see Relevant Websites Section) is an allied tool that can be used to design active site region, either to improve or modify ligand/substrate binding (Stiel *et al.*, 2016).

Conclusions

Protein design is an active, exciting area of research that has wide applications in drug design, medicine, and advancing the study of protein folding. When designing protein structures, one has to ask few questions and these questions drive or direct the use of tools to answer those questions.

1. Is the aim to engineer a protein's function/activity or designing a structure *de novo*?
2. Is the design fitting with the already known do's and don't's?
3. Is there another variable that is specific for the protein and its intended use?

There can be more additional questions added and the framework/checklist can act as a guide to the specific task in hand.

While, there are many success stories in *de novo* protein design, there are some challenges for the road ahead. Some of the challenges that have been discussed (Kuhlman *et al.*, 2009) are:

1. Sampling the conformational space of the backbone to move towards a completely flexible backbone based design. Currently, majority of the *de novo* designed proteins are on the basis of having a rigid scaffold or frame throughout the design pipeline.
2. To reduce the presence of alternate conformations from the desired/designed conformation.
3. Designing specific protein-protein interactions using a much easier and faster method.

Recently, Rocklin *et al.* from David baker's lab have used minimal proteins (proteins having a residue length below 50 amino acids) to understand the factors that determine protein folding. Specifically, they used computationally driven approach, where four topologies ($\alpha\alpha\alpha$, $\beta\alpha\beta\beta$, $\alpha\beta\beta\alpha$, and $\beta\beta\alpha\beta\beta$) were designed using 5000–40,000 sequences for each topology. Using yeast based proteolysis assay, a stability score was given to each designed protein, which enabled them to identify 2788 stable proteins. Such "massively-parallel" design has indeed pushed the limits of high-throughput design and the method can be applied to proteins of more than 50 amino acids in length (Rocklin *et al.*, 2017).

Acknowledgement

RMY acknowledges the infrastructural support given by Jaypee University of Information Technology, Waknaghat for completing this manuscript. Authorship was decided using a tic-tac-toe game with self. He also acknowledges Pulkit Anupam Srivastava for critical reading of the manuscript.

References

Baxa, M., Haddadian, E., Jumper, J., Freed, K., Sosnick, T., 2014. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. Proceedings of the National Academy of Sciences 111 (43), 15396–15401.

Bowie, J.U., Luthy, R., Eisenberg, D., 1991. A method to identify protein sequences that fold into a known three-dimensional structure. Science (New York, NY) 253 (5016), 164–170.

Brown, J.E., Klee, W.A., 1971. Helix-coil transition of the isolated amino terminus of ribonuclease. Biochemistry 10 (3), 470–476.

Crick, F.H.C., 1953. The packing of {α}-helices: Simple coiled-coils. Acta Crystallographica 6 (8–9), 689–697. doi:10.1107/S0365110X53001964.

Dahiyat, B.I., Mayo, S.L., 1997. De novo protein design: Fully automated sequence selection. Science (New York, NY) 278 (5335), 82–87.

Damborsky, J., Brezovsky, J., 2014. Computational tools for designing and engineering enzymes. Current Opinion in Chemical Biology 19, 8–16. doi:10.1016/j.cbpa.2013.12.003.

de Alba, E., *et al.*, 1996. Conformational investigation of designed short linear peptides able to fold into beta-hairpin structures in aqueous solution. Folding & Design 1 (2), 133–144.

de Alba, E., *et al.*, 1999. De novo design of a monomeric three-stranded antiparallel beta-sheet. Protein Science : A Publication of the Protein Society 8 (4), 854–865. doi:10.1110/ps.8.4.854.

DeGrado, W.F., Wasserman, Z.R., Lear, J.D., 1989. Protein design, a minimalist approach. Science (New York, NY) 243 (4891), 622–628.

Der, B.S., Kuhlman, B., 2011. Biochemistry. From computational design to a protein that binds. Science (New York, NY) 332 (6031), 801–802. doi:10.1126/science.1207082.

DiMaio, F., *et al.*, 2011. Modeling symmetric macromolecular structures in Rosetta3. PLOS ONE 6 (6), e20450. doi:10.1371/journal.pone.0020450.

Fazelinia, H., Cirino, P.C., Maranas, C.D., 2007. Extending Iterative Protein Redesign and Optimization (IPRO) in protein library design for ligand specificity. Biophysical Journal 92 (6), 2120–2130. doi:10.1529/biophysj.106.096016.

Gainza, P., *et al.*, 2013. OSPREY: Protein design with ensembles, flexibility, and provable algorithms. Methods in Enzymology 523, 87–107. doi:10.1016/B978-0-12-394292-0.00005-9.

Griffiths-Jones, S.R., Searle, M.S., 2000. Structure, folding, and energetics of cooperative interactions between the β-Strands of a de novo designed three-stranded antiparallel β-sheet peptide. Journal of the American Chemical Society 122 (35), 8350–8356. doi:10.1021/ja000787t.

Huyghues-Despointes, B.M., Scholtz, J.M., Baldwin, R.L., 1993. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. Protein Science: A Publication of the Protein Society 2 (10), 1604–1611. doi:10.1002/pro.5560021006.

Jiang, L., *et al.*, 2008. De novo computational design of retro-aldol enzymes. Science (New York, NY) 319 (5868), 1387–1391. doi:10.1126/science.1152692.

Kim, P.S., Baldwin, R.L., 1984. A helix stop signal in the isolated S-peptide of ribonuclease A. Nature 307 (5949), 329–334.

Kortemme, T., Ramirez-Alvarado, M., Serrano, L., 1998. Design of a 20-amino acid, three-stranded beta-sheet protein. Science (New York, NY) 281 (5374), 253–256.

Kuhlman, B., *et al.*, 2003. Design of a novel globular protein fold with atomic-level accuracy. Science (New York, NY) 302 (5649), 1364–1368. doi:10.1126/science.10894270.

Kuhlman, B., Jung Choi, E., Guntas, G., 2009. Future challenges of computational protein design. In: Protein Engineering and Design, CRC Press. Available at: http://dx.doi.org/10.1201/9781420076592.ch18.

Lacroix, E., Viguera, A.R., Serrano, L., 1998. Elucidating the folding problem of alpha-helices: Local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. Journal of Molecular Biology 284 (1), 173–191. doi:10.1006/jmbi.1998.2145.

Leitner, A., Walzthoeni, T., Kahraman, A., *et al.*, 2010. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. Molecular & Cellular Proteomics 9 (8), 1634–1649.

Lyskov, S., *et al.*, 2013. Serverification of molecular modeling applications: The rosetta online server that includes everyone (ROSIE). PLOS ONE 8 (5), e63906. doi:10.1371/journal.pone.0063906.

Lyu, P.C., *et al.*, 1990. Side chain contributions to the stability of alpha-helical structure in peptides. Science (New York, NY) 250 (4981), 669–673.

Malisi, C., Kohlbacher, O., Hocker, B., 2009. Automated scaffold selection for enzyme design. Proteins 77 (1), 74–83. doi:10.1002/prot.22418.

Mandell, D.J., Kortemme, T., 2009. Backbone flexibility in computational protein design. Current Opinion in Biotechnology 20 (4), 420–428. doi:10.1016/j.copbio.2009.07.006.

Marqusee, S., Baldwin, R.L., 1987. Helix stabilization by Glu-.Lys+ salt bridges in short peptides of de novo design. Proceedings of the National Academy of Sciences of the United States of America. 8898–8902.

Marqusee, S., Robbins, V.H., Baldwin, R.L., 1989. Unusually stable helix formation in short alanine-based peptides. Proceedings of the National Academy of Sciences of the United States of America 86 (14), 5286–5290.

Mitra, P., Shultis, D., Zhang, Y., 2013. EvoDesign: de novo protein design based on structural and evolutionary profiles. Nucleic Acids Research 41, W273–W280. doi:10.1093/nar/gkt384.

Munoz, V., Serrano, L., 1994. Elucidating the folding problem of helical peptides using empirical parameters. Nature Structural Biology 1 (6), 399–409.

Munoz, V., Serrano, L., 1995a. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. Journal of Molecular Biology 245 (3), 275–296.

Munoz, V., Serrano, L., 1995b. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. Journal of Molecular Biology 245 (3), 297–308. doi:10.1006/jmbi.1994.0024.

Munoz, V., Serrano, L., 1997. Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. Biopolymers 41 (5), 495–509. doi:10.1002/(SICI)1097-0282(19970415)41:5<495::AID-BIP2>3.0.CO;2-H.

Musil, M., *et al.*, 2017. FireProt: Web server for automated design of thermostable proteins. Nucleic Acids Research. doi:10.1093/nar/gkx285.

Nicholson, H., Becktel, W.J., Matthews, B.W., 1988. Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. Nature 336 (6200), 651–656. doi:10.1038/336651a0.

Panigrahi, P., *et al.*, 2015. Engineering proteins for thermostability with iRDP web server. PLOS ONE 10 (10), e0139486. doi:10.1371/journal.pone.0139486.

Park, S.H., Shalongo, W., Stellwagen, E., 1993. Residue helix parameters obtained from dichroic analysis of peptides of defined sequence. Biochemistry 32 (27), 7048–7053.

Pauling, L., Corey, R.B., 1953. Compound helical configurations of polypeptide chains: Structure of proteins of the alpha-keratin type. Nature 171 (4341), 59–61.

Ramirez-Alvarado, M., Blanco, F.J., Serrano, L., 1996. De novo design and structural analysis of a model beta-hairpin peptide system. Nature Structural Biology 3 (7), 604–612.

Raveh, B., London, N., Schueler-Furman, O., 2010. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins 78 (9), 2029–2040. doi:10.1002/prot.22716.

Richardson, J.S., 1981. The anatomy and taxonomy of protein structure. Advances in Protein Chemistry 34, 167–339.

Richardson, J.S., Richardson, D.C., 1988. Amino acid preferences for specific locations at the ends of alpha helices. Science (New York, NY) 240 (4859), 1648–1652.

Richardson, J.S., Richardson, D.C., 1989. The de novo design of protein structures. Trends in Biochemical Sciences 14 (7), 304–309.

Rocklin, G.J., *et al.*, 2017. Global analysis of protein folding using massively parallel design, synthesis, and testing. Science (New York, NY) 357 (6347), 168–175. doi:10.1126/science.aan0693.

Saraf, M.C., *et al.*, 2006. IPRO: An iterative computational protein library redesign and optimization procedure. Biophysical Journal 90 (11), 4167–4180. doi:10.1529/biophysj.105.079277.

Schafmeister, C.E., *et al.*, 1997. A designed four helix bundle protein with native-like structure. Nature Structural Biology 4 (12), 1039–1046.

Schenck, H.L., Gellman, S.H., 1998. Use of a designed triple-stranded antiparallel beta-Sheet to probe beta-sheet cooperativity in aqueous solution. Journal of the American Chemical Society 120 (11), 4869–4870. doi:10.1021/ja973984 + .

Scholtz, J.M., *et al.*, 1993. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. Biochemistry 32 (37), 9668–9676.

Sharman, G.J., Searle, M.S., 1998. Cooperative interaction between the three strands of a designed antiparallel β-sheet. Journal of the American Chemical Society 120 (21), 5291–5300. doi:10.1021/ja9705405.

Shoemaker, K.R., *et al.*, 1987. Tests of the helix dipole model for stabilization of alpha-helices. Nature 326 (6113), 563–567. doi:10.1038/326563a0.

Siegel, J.B., *et al.*, 2010. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science (New York, NY) 329 (5989), 309–313. doi:10.1126/science.1190239.

Simons, K.T., *et al.*, 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. Journal of Molecular Biology 268 (1), 209–225. doi:10.1006/jmbi.1997.0959.

Simons, K.T., *et al.*, 1999. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 34 (1), 82–95.

Sinz, A., Arlt, C., Chorev, D., Sharon, M., 2015. Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. Protein Science 24 (8), 1193–1209.

Smadbeck, J., *et al.*, 2013. Protein WISDOM: A workbench for in silico de novo design of biomolecules. Journal of Visualized Experiments: JoVE. 77). doi:10.3791/50476.

Stanger, H.E., Gellman, S.H., 1998. Rules for antiparallel β-sheet design: D-pro-gly is superior to L-Asn-Gly for β-hairpin nucleation. Journal of the American Chemical Society 120 (17), 4236–4237. doi:10.1021/JA973704Q.

Stiel, A.C., Nellen, M., Hocker, B., 2016. PocketOptimizer and the design of ligand binding sites. Methods in Molecular Biology Clifton NJ 1414, 63–75. doi:10.1007/978-1-4939-3569-7_5.

Struthers, M.D., Cheng, R.P., Imperiali, B., 1996. Design of a monomeric 23-residue polypeptide with defined tertiary structure. Science (New York, NY) 271 (5247), 342–345.

Wetzel, R., 1987. Harnessing disulfide bonds using protein engineering. Trends in Biochemical Sciences 12 (Suppl.), 478–482. Available at: https://doi.org/10.1016/0968-0004(87)90234-9.

Wood, C.W., *et al.*, 2017. ISAMBARD: An open-source computational environment for biomolecular analysis, modelling and design. Bioinformatics (Oxford). doi:10.1093/bioinformatics/btx352.
Xiong, H., *et al.*, 1995. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. Proceedings of the National Academy of Sciences of the United States of America 92 (14), 6349–6353.

Further Reading

Baldwin, R.L., 1995. Alpha-helix formation by peptides of defined sequence. Biophysical Chemistry 55 (1–2), 127–135.
Barlow, D.J., Thornton, J.M., 1988. Helix geometry in proteins. Journal of Molecular Biology 201 (3), 601–619.
Blanco, F., Ramirez-Alvarado, M., Serrano, L., 1998. Formation and stability of beta-hairpin structures in polypeptides. Current Opinion in Structural Biology 8 (1), 107–111.
Carey, P., 2008. Protein Engineering and Design. San Diego, CA: Academic Press.
Chakrabartty, A., Baldwin, R.L., 1995. Stability of alpha-helices. Advances in Protein Chemistry 46, 141–176.
Fisk, J.D., Gellman, S.H., 2001. A parallel beta-sheet model system that folds in water. Journal of the American Chemical Society. 343–344.
Guerois, R., De la Paz, M., 2006. Protein Design Methods in Molecular Biology, 340. Springer.
Jensen, K., 2010. Peptide and Protein Design for Biopharmaceutical Applications. Chichester: Wiley.
Lacroix, E., *et al.*, 1999. The design of linear peptides that fold as monomeric beta-sheet structures. Current Opinion in Structural 9 (4), 487–493.
Nowick, J.S., Cary, J.M., Tsai, J.H., 2001. A triply templated artificial beta-sheet. Journal of the American Chemical Society 123 (22), 5176–5180.
Park, S., Cochran, J., 2010. Protein Engineering and Design. Boca Raton: CRC Press.
Presta, L.G., Rose, G.D., 1988. Helix signals in proteins. Science (New York, NY) 240 (4859), 1632–1641.
Richardson, J.S., Richardson, D.C., 1988. Amino acid preferences for specific locations at the ends of alpha helices. Science New York, NY) 240 (4859), 1648–1652.
Rohl, C.A., Baldwin, R.L., 1998. Deciphering rules of helix stability in peptides. Methods in Enzymology 295, 1–26.
Scholtz, J.M., Baldwin, R.L., 1992. The mechanism of alpha-helix formation by peptides. Annual Review of Biophysics and Biomolecular Structure 21, 95–118. doi:10.1146/annurev.bb.21.060192.000523.
Searle, M.S., Ciani, B., 2004. Design of beta-sheet systems for understanding the thermodynamics and kinetics of protein folding. Current Opinion in Structural Biology 14 (4), 458–464. doi:10.1016/j.sbi.2004.06.001.
Serrano, L., 2000. The relationship between sequence and structure in elementary folding units. Advances in Protein Chemistry 53, 49–85.
Venkatraman, J., Shankaramma, S.C., Balaram, P., 2001. Design of folded peptides. Chemical Reviews 101 (10), 3131–3152.

Relevant Websites

http://agadir.crg.es/
Adagir.
http://www.cs.duke.edu/donaldlab/osprey.php
Donald lab.
http://www.grigoryanlab.org/drawcoil
Drawcoil 1.0.
http://rzlab.ucr.edu/scripts/wheel/wheel.cgi
Helical Wheel Projections - RZLab.
http://irdp.ncl.res.in
iRDP.
https://loschmidt.chemi.muni.cz/fireprot
LOSCHMIDT laboratories.
http://www.eb.tuebingen.mpg.de/research/research-groups/birte-hoecker/algorithms-and-software/pocketoptimizer.html
Max Planck Institute for Developmental Biology.
http://www.eb.tuebingen.mpg.de/research/research-groups/birte-hoecker/algorithms-and-software/scaffoldselection.html
Max Planck Institute for Developmental Biology.
http://maranas.che.psu.edu/submission/IPRO_2.htm
PENNSTATE.
http://emboss.bioinformatics.nl/cgi-bin/emboss/pepwheel
Pepwheel.
http://atlas.princeton.edu/proteinwisdom
Protein WISDOM.
https://pypi.python.org/pypi/isambard
Python.
https://www.rosettacommons.org/
Rosetta commons.
http://rosettadesign.med.unc.edu
Rosetta commons.
http://rosie.rosettacommons.org/
Rosetta commons.
https://zhanglab.ccmb.med.umich.edu/EvoDesign
Zhang lab.

Biographical Sketch

Ragothaman Yennamalli is an Assistant Professor at Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India. He completed his PhD in Computational Biology and Bioinformatics from Jawaharlal Nehru University, New Delhi. He conducted postdoctoral research at Jawaharlal Nehru University in India and at Iowa State University (2009–2011), University of Wisconsin-Madison (2011–2012), and Rice University (2012–2014) in USA. He is a structural and computational biologist with more than a decade of experience in predictive modelling and biomolecular simulation projects. Dr. Yennamalli's skills involve molecular docking, molecular dynamics simulation, coarse grained modelling, machine learning, data mining, data analytics, and molecular modelling. Website: http://bit.ly/raghu_juit.