Contents

1 Setting protein engineering goals

Summary. Protein engineering is the modification of proteins to improve their properties, usually by replacing amino acid residues with other amino acids. Proteins evolved in nature for their biological role, but for other applications, they usually require improvements. The four types of improvement goals are changes in stability, binding, reactivity and selectivity. Biotechnology uses engineered proteins in three main application areas: medicine, agriculture, and industry. Protein engineering in medicine has tuned the bioavailability of insulin and increased the potency and plasma half-life of monoclonal antibodies. Protein engineering in agriculture has increased the insect resistance of crops and stabilized enzymes for use as feed additives. Protein engineering for industrial applications include stabilizing detergent enzymes and engineering enzymes for pharmaceutical synthesis. Replacing chemical methods for pharmaceutical manufacture with biocatalysis usually yields a greener, more environmentally friendly process.

Key learning goals

- While evolution in nature improves proteins for their natural function, application of proteins to non-natural functions requires protein engineering.
- The four main goals of protein engineering are improvements in stability, binding, reactivity and selectivity. Identifying the goal is a critical first step of any protein engineering project.
- Engineered proteins have applications in medicine, agriculture and industry. Your home likely contains engineered proteins.

1.1 Introduction

Over millennia humans have domesticated crops and animals by breeding to select for desired traits. Current wheat produces more grain than wild wheat; dogs are better com-

panions than wolves. The molecular understanding of biology in recent decades created more powerful approaches to improve biology for human applications. These applications of biology to make useful products in medicine, agriculture and industry are collectively called biotechnology. The worldwide sales of biotechnology were ~US$160 billion in 2017.[1] The most significant application area is medicine, sometimes called red biotechnology, Figure 1.1. The next largest is agriculture, called green biotechnology and the smallest is industry, called white biotechnology.

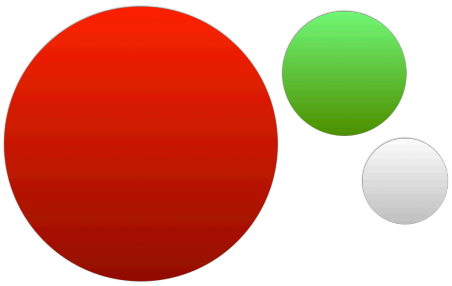

Figure 1.1. Biotechnology includes applications in medicine (red), agriculture (green) and industry (white). The areas of the circles approximate the relative worldwide sales in each area.

The two primary technologies within biotechnology are genetic engineering and protein engineering. Genetic engineering changes the genetic code of an organism. The most common genetic engineering is adding a gene from one organism to another organism so that it makes a protein that it did not make previously. Such recombinant organisms gain new abilities. For example, adding the gene encoding human insulin to *E. coli* bacteria allowed manufacture of this essential therapeutic protein. Adding genes for proteins toxic to insects into crop plants made them resistant to insects. Combining genes to degrade different hydrocarbons into one bacteria enabled it to degrade oil spills more effectively.

Protein engineering is the modification of proteins to improve them for specific applications. Altering the gene that encodes the natural protein creates a new gene that encodes an improved variant protein. Changing the protein can improve its stability, binding, catalytic activity or even give it new functions.[2] The modifications are typically substitution of one amino acid residue with another one, but sometimes include adding or removing amino acid residues, even entire domains of a protein. For example, engineering of subtilisin, a detergent protease, replaced an easily oxidized amino acid with one less easily oxidized, thereby increasing its stability to bleach during washing,[3] Figure 1.2.

Applications that differ from the natural function usually require protein engineering Proteins have evolved in nature for specific natural functions. When proteins are used for another function or in a different way, they will likely require engineering. Natural selection chooses proteins that improve fitness in nature, not for various applications that humans identify. Natural proteins may be accidentally suitable for these applications, but won't be optimized for them. For example, nature has not evolved proteins

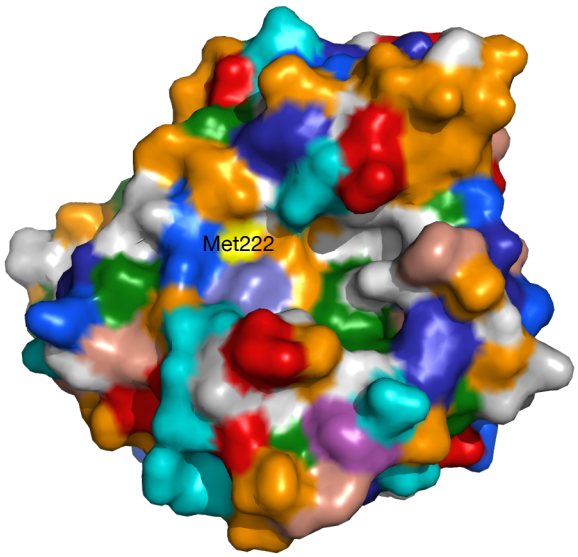

Figure 1.2. Surface representation of the protease subtilisin where the twenty amino acids are colored in one of twelve different colors according to traditional amino acid properties. Protein engineering increases the stability of this protease to bleach by replacing methionine 222 (labelled yellow patch near the center) with alanine.

tolerant to organics solvents since no organisms live in organic solvents, but some proteins evolved to tolerate high temperatures often also tolerate organic solvents. In some cases, the different requirements are obvious. For example, an enzyme-catalyzed manufacture of a pharmaceutical intermediate may involve an unnatural substrate, organic cosolvents and elevated temperatures. In these cases, protein engineering can adjust the active site to fit the new substrate to increase reaction rate and stabilize the protein to organic solvents and elevated temperatures.

In other cases, the differences between the natural protein function and application may be subtle. For example, human insulin regulates the glucose levels in the blood. Patients with type I diabetes cannot make insulin and require injections of human insulin. Although the goal is the same - to maintain glucose levels, the dosing method differs. Patients inject insulin several times per day in contrast to the pancreas continuously secreting varying amounts of insulin. This difference in dosing required engineering fast-acting variants of human insulin for injection before meals and slow-acting variants to maintain glucose levels overnight.

Evolution does not maximize the properties of a protein, but only maintains the minimum necessary. Natural selection cannot distinguish between a protein that is improved just enough so that another protein limits function and a protein that is much better. Selection favors both equally since the much better property does not contribute to fitness. A hammer made from extra-hard steel does not drive nails any better than a hammer with steel just hard enough.

An example of 'just enough' is that most proteins are just stable enough for their natural function.[4],[5] This marginal stability is not an inherent limitation of proteins since single amino acid substitutions can stabilize most proteins without compromising activity.[6] The reason for the marginal stability is evolution cannot select for more than the minimum needed stability.[7] One can also generalize this 'just enough' quality to other protein properties. Enzymes are just fast enough for their function, they bind their substrates just tightly enough and discriminate between substrates just enough for them to function.

1.2 Four categories of protein engineering goals

> To the person who does not know where he wants to go there is no favorable wind. — Seneca

Although proteins can be engineered to be better for many different types of applications, the improvements themselves fall into four categories: stability, binding, reactivity or selectivity, Table 1.1. Stability refers to the ability of a protein to maintain its function, usually binding to a target or catalyzing a reaction. Increasing protein stability allows it to tolerate harsh conditions such as high temperatures or the presence of bleach as in the detergent protease example above. It may make the protein last longer under normal application or storage conditions. For biopharmaceuticals, stability may also refer to extending the serum lifetime of the protein. Most proteins are rapidly removed from the blood thus limiting their therapeutic benefit. Increasing the serum lifetime enhances the therapeutic benefit.

Table 1.1. Four categories of goals in protein engineering.

Goal	Examples of protein improvements
stability	- tolerate heat, organic solvents, extremes of pH or other harsh conditions - tolerates storage or use for a long time at normal conditions - tolerate destabilizing substitutions - longer serum lifetime - tolerate digestive proteases
binding	- bind target ligand more tightly or less tightly - avoid binding to antibodies to avoid allergic reaction
catalytic activity	- faster catalysis of existing substrates - expand catalysis to new substrates - inactivate undesired catalytic activity - enable catalysis of a new chemical reaction
selectivity	- favor binding one ligand over another in the same solution - favor reaction of one of several competing substrates - favor formation of one of several possible products

Improvements in binding increase the affinity of a protein for its target. For biopharmaceuticals, it is also important that they do not bind to human antibodies so that they do not cause an allergic reaction. Humanization of proteins engineers them so that they do not bind to human antibodies. Improvements in reactivity can increase reaction rates to existing substrates or create new ability by expanding reactivity to new substrates or even new reaction types. Selectivity involves differences in binding or reactivity for competing molecules in the same solution. Selectivity may refer to the ability of a protein to bind one of several ligands or to catalyze reaction of one of several substrates. It can also refer to the selective formation of one of several possible products from a single substrate.

Identifying the protein engineering goal is a critical first step in any protein engineering project because it defines the approaches that one can use. Ideally one also defines the amount of change needed because it reveals how difficult the engineering will be.

Sometimes protein engineering changes several protein properties so that it can be hard to identify which property is the true protein engineering goal. For example, researchers wanted to engineer a protease that could degrade the protein gluten in the human digestive track so they could add it to the diet of gluten-intolerant patients.[8] The researchers started with the protease KumaWT, which favored hydrolysis after a ProArg or ProLys sequence and wanted to change it to favor hydrolysis after a ProGln sequence since gluten contains many ProGln amino acid pairs in its sequence. The researchers mentioned "enhance its activity" suggesting the goal was increased reactivity. They also mentioned "desired oligopeptide specificity" suggesting that increasing the selectivity

for ProGln-containing peptides over other peptides might also be important. Finally, they mentioned "redesign the S1 binding pocket" suggesting that increased binding of the ProGln peptides might also be important. Indeed, the results showed that all three of these properties - reactivity, selectivity and binding - changed. The changes introduced into KumaWT increased the reactivity toward ProGln-containing peptides, increased the selectivity of the enzyme for ProGln versus ProArg sequences and increased the binding of the ProGln sequence to the protease binding site.

In reality only one of these properties was the true goal. It is the property that *must* change in order to yield an improved protein. The other changes are side effects of their engineering to achieve this goal. First, imagine if the catalytic activity toward the ProGln sequence increased, but there was no change in the selectivity (the ProLys sequence also reacted) and no change in the binding (the increase in reactivity was not associated with better binding). In this case the engineered protein would still be improved because it would degrade the ProGln containing peptides. Increasing reactivity is the true goal. Second, imagine if selectivity improved without changing the catalytic activity. (A decrease in the activity toward the ProLys sequence could increase the selectivity for the ProGln sequence without increasing the activity toward the ProGln sequence.) Similarly, imagine that the binding of the ProGln sequence to the enzyme remained unchanged. In this case the enzyme has not improved. No more ProGln would be cleaved than with the starting enzyme. Thus, selectivity is not the true goal and the increase is selectivity is a side effect of increased reactivity. Finally, consider the case where the binding improved, but not the selectivity or reactivity. Again, without an increase in reactivity, no more ProGln sequences would be cleaved. The authors hypothesized that the reason for the poor reactivity was that the ProGln sequence binds poorly. Thus, increasing bindin was an approach to increase reactivity. If they achieved increased binding without an increase in reactivity, they would have failed and tried another approach to increase reactivity. Increased binding alone would not have yielded a better enzyme. Thus, protein engineering can simultaneously change several protein properties. The true protein engineering goal is the property that *must* change in order to see an improvement in the proposed application.

1.3 Engineering proteins for medicine

Biopharmaceuticals are biomolecules used as drugs. They can be proteins such as antibodies or enzymes; they can also be nucleic acids such as micro RNAs and even complete microbes such as attenuated live vaccines. This text considers only protein biopharmaceuticals.

The first protein biopharmaceuticals were recombinant equivalents of the natural protein where the biotechnology advantage was the ability to produce large amounts of protein. For example, blood factors that were isolated from donated blood could now be produced in mammalian cell cultures due to genetic engineering. The subsequent generations of blood factors were improved engineered variants. For example, engineering increased the potency of the blood coagulation protein, Factor VII used to treat hemophilia, so that each dose required less protein,[9] Table 1.2. The action of this blood coagulation protein depends on its binding to a membrane surface containing acidic

Table 1.2. Examples of protein engineering to improve biopharmaceuticals.

Category	Example of improvement	Property change
blood factors	more potent coagulation factor VII	- increase **binding** to damaged cells - increased **reactivity** (proteolysis)
thrombolytics & anticoagulants	more potent and longer plasma half-life of tissue plasminogen activator	- increase **binding** to fibrin - reduce **binding** to receptor that removes proteins from blood
hormones	fast-acting or long-lasting insulins	- alter **binding** with other monomers
monoclonal antibodies	reduced immunogenicity of mouse-derived antibody	- reduce **binding** to immune response proteins
growth factors	longer serum half-life of granulocyte colony-stimulating factor	- increase **binding** to receptors that recycle proteins to blood stream
vaccines	less toxic pertussis (whooping cough) vaccine	- inactivate **reactivity** that causes toxicity
other	less immunogenic asparaginase	- reduce **binding** to immune response proteins

phospholipids, which is characteristic of damaged vascular cells or of platelets that adhere to the damage as the clot forms. Mutagenesis of the membrane-binding domain of Factor VII yielded variants that bound more tightly to the membrane and therefore were more potent. The best variant contained five substitutions and was 149-296 fold better than wild-type Factor VII. Other researchers increased its potency with three amino acid substitutions that increased its catalytic activity 30-fold.[10] Proteolytic activity of Factor VII activates the next factor in the coagulation cascade.

Insulin variants engineered for altered bioavailability are even better than human insulin at controlling blood sugar levels. The engineering relied on the fact that monomeric insulin is the active form, while multimeric forms are inactive. Insulin lispro is a fast-acting insulin analog for injections before meals because it favors the monomeric form.[11] The amino acid sequence of lispro differs only in the reversal of the penultimate lysine and proline residues on the C-terminal end of the B-chain. (Insulin consists of 51 amino acids arranged as a dimer of an A-chain and a B-chain linked by disulfide bonds.) This residue change blocks the formation of insulin dimers and hexamers and increases the amount of monomeric insulin (the active form), thereby

creating a faster-acting variant. In contrast, substitutions to create a slow-acting insulin analog, glargine, promote association of the monomers into an insoluble precipitate, which then slowly releases monomeric insulin to maintain basal levels of insulin for an extended time.

An essential class of therapeutic proteins is monoclonal antibodies and monoclonal antibody drug conjugates. Common goals of antibody engineering are increasing binding the target and minimizing immunogenicity.[12] One example of a therapeutic monoclonal antibody is trastuzumab (Herceptin®) used to treat breast cancer. This antibody interferes with the HER2 receptor (human ErbB2 receptor tyrosine kinase), which signals cell proliferation. Cancer cells, usually breast cancer, overexpress the HER2 receptor causing the cancer cells to grow uncontrollably. Interfering with the receptor stops the uncontrolled growth.

Trastuzumab is a humanized version of a mouse antibody, Figure 1.3. Humanization prevents an allergic reaction to a mouse antibody. This humanization involved first, the transplantation of the complementarity-determining regions from the mouse antibody into a human IgG antibody and second, an optimization of the surrounding regions, called framework regions.[13] The transplantation of the complementarity-determining regions changed 24 amino acids in the light chain and 32 in the heavy chain. This substitution did not yield a useful antibody. The engineered version bound less tightly than the mouse antibody and did not block cancer cell proliferation upon binding. Next, an additional seven substitutions in the surrounding (framework) regions increased the binding to the target 100-fold making it three-fold tighter than the mouse antibody. This tighter binding restored the ability of the antibody to block cell proliferation. Trastuzumab was approved in the USA in 1998 to treat metastatic breast cancer.

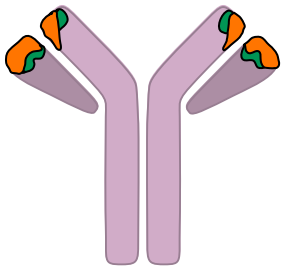

Figure 1.3. To engineer trastuzumab, researchers combined the complementarity-determining regions from a mouse antibody (orange; 56 amino acids) with a human IgG antibody (lavender). Full activity required seven additional substitutions (green) outside the binding region.

Growth factors and cytokines, such as vascular endothelial growth factor, epidermal growth factor and granulocyte colony-stimulating factor, stimulate cell recruitment, proliferation, morphogenesis, and differentiation. Inhibitors of these growth factors act as anticancer drugs, while enhanced growth factors promote wound repair. Natural growth factors lack stability and specificity; they are prone to degradation in serum and have multiple activities. In one example of protein engineering of a cytokine, Sarkar

and coworkers increased the half-life and potency of granulocyte colony-stimulating factor by modifications that increased its recycling back to the bloodstream.[14] Cantor and coworkers engineered asparaginase, an enzyme for leukemia treatment, to be less immunogenic.[15] Treatment of childhood acute lymphoblastic leukemia involves injections of asparaginase, which cleaves asparagine to aspartate, thereby depriving the cancerous cells of asparagine. Healthy cells can make asparagine, but these cancer cells cannot. Humans do not have an asparaginase enzyme, so the treatment uses recombinant enzyme from Escherichia coli. One limitation is that some patients' immune systems recognize this enzyme as a non-human protein and inhibit it by binding it with an antibody or even induce a severe allergic reaction. Cantor and coworkers replaced the amino acid sequences on the surface of *E. coli* asparaginase that are likely to be recognized by human T-cells with non-recognized sequences. The replacements maintained the catalytic activity of the asparaginase. Vaccines against pertussis (whooping cough) contain engineered proteins. The bacteria that causes pertussis, Bordetella pertussis, secrete a five-subunit protein, known as pertussis toxin. Immunization with the pertussis toxin protein or just its S1 subunit protects against disease. However, the S1 subunit is an enzyme that catalyzes the ADP-ribosylation of GTP-binding proteins, thereby causing toxicity. Two amino acid substitutions in the S1 subunit - Glu129Gly, Arg9Lys - deactivated its catalytic activity, and therefore toxicity, while retaining its immunogenic potential).[16]

The mRNA vaccines against the SARS-CoV-2 virus encode the spike protein, which sits on the virus surface and binds to cell surface proteins during infection. Upon binding to the target cell, the spike protein changes conformation. Researchers reasoned that spike protein in the pre-binding conformation would make a better vaccine that the spike protein in the post-binding conformation. Binding of antibodies to the spike protein before it binds to the cell could prevent infection. To stabilize the pre-binding conformation, researchers replaced six residues in the spike protein with proline.[17] Proline limits the flexibility at those sites due to its ring structure and keeps the spike protein in the pre-binding conformation.

1.4 Engineering proteins for agriculture

Agricultural biotechnology aims mainly to improve productivity of crops and animals. Some products, like bovine somatotrophin (a growth hormone), are copies of the natural protein produced in recombinant E. coli bacteria. The pituitary gland produces small amounts of somatotrophin, but production in bacteria increases the availability and lowers the cost. Treating cows with somatotrophin increases milk production. In other cases, natural proteins are improved by engineering, Table 1.3.

Glyphosate is a broad-spectrum, systemic herbicide that acts by inhibiting the enzyme EPSP synthase in aromatic amino acid biosynthesis. Farmers can control weeds by spraying glyphosate on fields, but only if the crops are insensitive to glyphosate. Adding a glyphosate-insensitive variant of EPSP synthase from Agrobacterium created the first glyphosate-resistant soybean plants.[18] The next generation of glyphosate-resistant crops contain enzymes such as oxidases or acetyl transferases, which inactivate glyphosate. Protein engineering to increase the catalytic activity of these enzymes is an

Table 1.3. Examples of protein engineering to improve agriculture.

Category	Example of improvement	Property change
herbicide-resistant crops	glyphosate-resistant soybeans	- add new **reactivity** (glyphosate oxidase)
insect-resistant crops	cotton resistant to bollworm	- increase **binding** of *Bt* toxins to target proteins in insect gut
enhanced photosynthesis	faster growth of *Arabidopsis* at higher temperatures	- increase **stability** of rubisco activase
modified products	algae producing an increased fraction of medium-chain fatty acids	- increased **reactivity** toward medium-chain fatty acid precursors
feed additives	animal feed with higher phosphorus availability	- increase **stability** of phytase

important goal.[19]

Genetic modification of crops such as cotton, corn, and rice to express insecticidal crystal proteins produced by *Bacillus thuringiensis* protects the crops against insect pests. The toxicity of these proteins varies toward different insects, and some insects develop resistance to these toxins. Enhancing the toxicity of these proteins by increasing the their binding to target proteins in the insect gut is an essential goal in agricultural biotechnology.[20]

Heat stress hinders plant growth and lowers crop yields mainly due to reduced photosynthesis. This reduction is mainly due to an inactivation of a thermolabile ATPase, Rubisco activase, whose role is to maintain ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) in its active state. Kurek and coworkers engineered a thermostable Rubisco activase, which improved the growth of Arabidopsis plants at higher temperatures.[21]

Whittle and Shanklin altered the chain length selectivity of a plant fatty acid desaturase to accept 16-carbon instead of 18-carbon substrate.[22] Lin and Lee engineered algae to produce more medium chain length fatty acids relative to long-chain-length fatty acids for use in biodiesel production.[23]

Engineering a more stable feed additive enzyme increases the nutritional value of feed and reduces pollution. Phosphorus in grains and oilseeds often occurs as phytic acid (*myo*-inositol hexakisphosphate). Non-ruminants such as pigs and chicken digest phytic acid incompletely leading to high phosphorus in wastewater. The enzyme phytase in animal feed aids digestion of phytic acid by catalyzing its hydrolysis. Forming animal

Table 1.4. Examples of protein engineering to improve industrial enzymes.

Category	Example of improvement	Property change
detergents	bleach-tolerant protease	- **stabilize** subtilsin to oxidation
food	simplify starch hydrolysis to maltodextrins and glucose	- **stabilize** α-amylase to low pH
other (e.g., biofuels, pharmaceuticals)	engineer transaminase for manufacture of diabetes drug	- increase **reactivity** toward target substrate - increase **stability** to reaction conditions

feed pellets requires pressing at high temperatures, so the phytases have been engineered for higher thermal stability to withstand this step.[24]

1.5 Engineering proteins for industry

Industry uses enzymes for manufacturing and similar non-natural applications, Table 1.4. The three main application areas are detergents, food and beverage, and other, which includes enzymes for biofuel and pharmaceutical synthesis. The advantages of using enzymes over chemical reagents or catalysts are that they are faster, greener and more selective. The primary engineering goals are stability and faster catalysis, both of which lower the cost of the enzyme.

Detergent enzymes such as proteases speed removal of food, blood and other stains on clothing. Subtilisin was the first industrial enzyme to be engineered.[3] Subtilisin tolerated hot water and surfactants, but bleach rapidly inactivated it, which limited its usefulness. Replacement of an oxidation-sensitive methionine 222 in the active site (Figure 1.2) by alanine stabilized subtilisin while maintaining high activity. The cost of these industrial proteins typically $100/kg. In contrast, proteins for medical applications can cost 10^5-fold more, $10,000/g, because their manufacture and regulatory approval are more complicated.

The largest volume application in food and beverages is the conversion of cornstarch to glucose catalyzed by α-amylase and glucoamylase. Glucose isomerase can isomerize the resulting glucose to high fructose corn syrup, which tastes sweeter than glucose. This process uses high temperatures to prevent microbial growth and high concentrations to minimize reactor size. Both α-amylase[25] and glucose isomerase[26] have been engineered to tolerate high temperatures and high glucose concentration. Glucose isomerase has a half-life = 50-100 d at the operating temperature of 55 °C. More than 500 tons of glucose isomerase produce $\approx 10^x$ tons of high fructose corn syrup each year, corresponding to productivities of >10,000 kg corn syrup per kg enzyme.

The use of enzymes for chemical synthesis is known as biocatalysis.[27] In pharmaceuti-

cal manufacture, the most significant application is the penicillin G amidase- catalyzed manufacture penicillin and cephalosporin antibiotics. This process yields more than 10,000 tons of antibiotics annually with a productivity of >600 kgs product/kg enzyme.[vanderwielenProcessTechnologyProcess2001] These antibiotics are natural products, so they are the natural substrates for this enzymes. However, many other pharmaceuticals are not natural products, so the ability of enzymes to act on them is accidental. Engineering of enzymes to improve their ability to act on unnatural substrates and tolerate the harsh conditions of a chemical reactor is a common goal.

Engineered enzymes can also create new biochemical pathways within cells, and the whole cells can be used for synthesis. For example, Ran and Frost expanded the substrate range of an aldolase to create a new metabolic pathway to make shikimic acid for an influenza drug synthesis.[28] Keasling and coworkers combined enzymes from different organisms and biochemical pathways to create a new biochemical pathway for synthesis of artemisinin, an anti-malarial.[29]

1.5.1 Green chemistry

One reason to use enzymes for manufacture of pharmaceuticals is to decrease the environmental impact of the process. The risk associated with a chemical depends both on how dangerous it is (hazard) and on one's contact with it (exposure), eq. 1.1.

$$risk = hazard \cdot exposure \tag{1.1}$$

In the past, governments and industry focused on reducing risk by minimizing exposure. Lab coats, safety glasses, and other chemical handling rules limit the exposure of workers to hazardous chemicals. Green chemistry is hazard prevention approach.[30]. Preventing problems is inevitably easier and less expensive than contending with difficulties after they occur. Green chemistry is the design of chemical products and processes that reduce or eliminate hazardous substance. Replacing chemical manufacturing steps with enzyme-catalyzed steps is an essential tool in green chemistry. Enzyme-catalyzed steps often replace hazardous solvents with water and hazardous catalysts with biodegradable enzymes. One example of greener pharmaceutical manufacture is the improved synthesis of sitagliptin, the active ingredient in an oral type 2 diabetes drug. The difficult step is the addition of the amino group with the correct orientation, Figure 1.4.

Researchers at Merck had already dramatically improved the original synthesis of sitagliptin by using a catalytic asymmetric hydrogenation with a rhodium catalyst. This improved synthesis won a Green Chemistry award in 2006. Biocatalysis further improved this synthesis. Biocatalysis used a transaminase-catalyzed formation of an amine, which eliminated four steps including one that required a high pressure reactor.[31] It resulted in a 10-13% higher overall yield and 19% less total waste. This improved synthesis won a second Green Chemistry award in 2010. Enabling this synthesis required extensive protein engineering of the transaminase. The starting transaminase did not react at all with the required substrate and was unstable under the reaction conditions. The engineering replaced 27 amino acids creating a highly active, highly selective and stable enzyme.

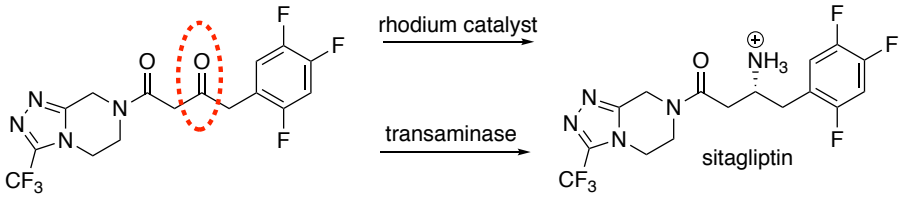

Figure 1.4. Two improved routes to sitagliptin showing the difficult step of inserting the amino group with the correct configuration. Improvements in the chemocatalytic process (rhodium catalyst) won a Green Chemistry award in 2006. Further improvements by switching to a biocatalytic process (transaminase) won a second Green Chemistry award in 2010.

1.6 Rational design and directed evolution

The two main strategies of protein engineering are rational design and directed evolution. Rational design predicts the changes needed to improve the protein based on the protein structure, molecular basis of its action and from computer modeling. In many cases, this approach yields dramatic improvements, but sometimes the needed changes are difficult to predict. In these cases, researchers use directed evolution. Directed evolution generates variants, screens these variants for improved properties and repeats the cycle. The first part of this text describes rational design approaches to protein engineering, while the second part describes directed evolution.

References

1. Morrison, C., & Lähteenmäki, R. (2018). Public biotech in 2017-the numbers. *Nat. Biotechnol.*, *36*(7), 576–584. https://doi.org/10.1038/nbt.4175
2. Brannigan, J. A., & Wilkinson, A. J. (2002). Protein engineering 20 years on. *Nat. Rev. Mol. Cell Biol.*, *3*(12), 964–970. https://doi.org/10.1038/nrm975
3. Estell, D. A., Graycar, T. P., & Wells, J. A. (1985). Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J. Biol. Chem.*, *260*(11), 6518–6521. http://www.jbc.org/content/260/11/6518.short
4. Pace, C. N., & Hermans, J. (1975). The stability of globular protein. *CRC Crit. Rev. Biochem.*, *3*(1), 1–43.
5. Taverna, D. M., & Goldstein, R. A. (2002). Why are proteins marginally stable? *Proteins*, *46*(1), 105–109. https://doi.org/10.1002/prot.10016
6. Fersht, A. R., & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Curr Opin Struct Biol*, *3*(1), 75–83. https://doi.org/10.1016/0959-440X(93)90205-Y
7. Bloom, J. D., Lu, Z., Chen, D., Raval, A., Venturelli, O. S., & Arnold, F. H. (2007). Evolution favors protein mutational robustness in sufficiently large populations. *BMC Biol.*, *5*(1), 97–21. https://doi.org/10.1186/1741-7007-5-29
8. Gordon, S. R., Stanley, E. J., Wolf, S., Toland, A., Wu, S. J., Hadidi, D., Mills, J. H., Baker, D., Pultz, I. S., & Siegel, J. B. (2012). Computational design of an α-gliadin peptidase. *J. Am. Chem. Soc.*, *134*(50), 20513–20520. https://doi.org/10.1021/ja30

94795
9. Harvey, S. B., Stone, M. D., Martinez, M. B., & Nelsestuen, G. L. (2003). Mutagenesis of the γ-carboxyglutamic acid domain of human factor VII to generate maximum enhancement of the membrane contact site. *J. Biol. Chem.*, *278*(10), 8363–8369. https://doi.org/10.1074/jbc.M211629200
10. Persson, E., Olsen, O. H., Bjørn, S. E., & Ezban, M. (2012). Vatreptacog alfa from conception to clinical proof of concept. *Semin. Thromb. Hemost.*, *38*(3), 274–281. https://doi.org/10.1055/s-0032-1302442
11. Bolli, G. B., Di Marchi, R. D., Park, G. D., Pramming, S., & Koivisto, V. A. (1999). Insulin analogues and their potential in the management of diabetes mellitus. *Diabetologia*, *42*(10), 1151–1167. https://doi.org/10.1007/s001250051286
12. Ryu, J. K., Kim, H. S., & Nam, D. H. (2012). Current status and perspectives of biopharmaceutical drugs. *Biotechnol. Bioprocess Eng.*, *17*(5), 900–911. https://doi.org/10.1007/s12257-012-0095-1
13. Carter, P., Presta, L., Gorman, C. M., Ridgway, J. B. B., Henner, D., Wong, W. L. T., Rowland, A. M., Kotts, C., Carver, M. E., & Shepard, H. M. (1992). Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.*, *89*(10), 4285–4289. http://www.pnas.org/content/89/10/4285.long
14. Sarkar, C. A., Lowenhaupt, K., Horan, T., Boone, T. C., Tidor, B., & Lauffenburger, D. A. (2002). Rational cytokine design for increased lifetime and enhanced potency using pH-activated "histidine switching". *Nat. Biotechnol.*, *20*(9), 908–913. https://doi.org/10.1038/nbt725
15. Cantor, J. R., Yoo, T. H., Dixit, A., Iverson, B. L., Forsthuber, T. G., & Georgiou, G. (2011). Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. *Proc. Natl. Acad. Sci. U. S. A.*, *108*(4), 1272–1277. https://doi.org/10.1073/pnas.1014739108
16. Loosmore, S. M., Zealey, G. R., Boux, H. A., Cockle, S. A., Radika, K., Fahim, R. E., Zobrist, G. J., Yacoob, R. K., Chong, P. C., & Yao, F. L. (1990). Engineering of genetically detoxified pertussis toxin analogs for development of a recombinant whooping cough vaccine. *Infect Immun*, *58*(11), 3653–3662. https://doi.org/10.1128/iai.58.11.3653-3662.1990
17. Hsieh, C.-L., Goldsmith, J. A., Schaub, J. M., DiVenere, A. M., Kuo, H.-C., Javanmardi, K., Le, K. C., Wrapp, D., Lee, A. G., Liu, Y., Chou, C.-W., Byrne, P. O., Hjorth, C. K., Johnson, N. V., Ludes-Meyers, J., Nguyen, A. W., Park, J., Wang, N., Amengor, D., … McLellan, J. S. (2020). Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*, *369*(6510), 1501–1505. https://doi.org/10.1126/science.abd0826
18. Padgette, S. R., Kolacz, K. H., Delannay, X., Re, D. B., LaVallee, B. J., Tinius, C. N., Rhodes, W. K., Otero, Y. I., Barry, G. F., Eichholtz, D. A., Peschke, V. M., Nida, D. L., Taylor, N. B., & Kishore, G. M. (1995). Development, identification, and characterization of a glyphosate-tolerant soybean line. *Crop Sci.*, *35*(5), 1451–1461. https://doi.org/10.2135/cropsci1995.0011183x003500050032x
19. Castle, L. A., Siehl, D. L., Gorton, R., Patten, P. A., Chen, Y. H., Bertain, S., Cho, H.-J., Duck, N., Wong, J., Liu, D., & Lassner, M. W. (2004). Discovery and directed evolution of a glyphosate tolerance gene. *Sci. Wash.*, *304*(5674), 1151–1154. https://doi.org/10.1126/science.1096770

20. Bravo, A., & Soberon, M. (2008). How to cope with insect resistance to Bt toxins? *Trends Biotechnol.*, *26*(10), 573–579. https://doi.org/10.1016/j.tibtech.2008.06.005
21. Kurek, I., Chang, T. K., Bertain, S. M., Madrigal, A., Liu, L., Lassner, M. W., & Zhu, G. (2007). Enhanced thermostability of Arabidopsis rubisco activase improves photosynthesis and growth rates under moderate heat stress. *Plant Cell*, *19*(10), 3230–3241. https://doi.org/10.1105/tpc.107.054171
22. Whittle, E., & Shanklin, J. (2001). Engineering δ^9 -16:0-acyl carrier protein (acp) desaturase specificity based on combinatorial saturation mutagenesis and logical redesign of the castor δ^9 -18:0-acp desaturase. *J. Biol. Chem.*, *276*(24), 21500–21505. https://doi.org/10.1074/jbc.M102129200
23. Lin, H., & Lee, Y. K. (2017). Genetic engineering of medium-chain-length fatty acid synthesis in Dunaliella tertiolecta for improved biodiesel production. *J Appl Phycol*, *29*(6), 2811–2819. https://doi.org/10.1007/s10811-017-1210-7
24. Yao, M. Z., Zhang, Y. H., Lu, W. L., Hu, M. Q., Wang, W., & Liang, A. H. (2012). Phytases: Crystal structures, protein engineering and potential biotechnological applications. *J. Appl. Microbiol.*, *112*(1), 1–14. https://doi.org/10.1111/j.1365-2672.2011.05181.x
25. Richardson, T. H., Tan, X., Frey, G., Callen, W., Cabell, M., Lam, D., Macomber, J., Short, J. M., Robertson, D. E., & Miller, C. (2002). A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable α-amylase. *J. Biol. Chem.*, *277*(29), 26501–26507. https://doi.org/10.1074/jbc.M203183200
26. Quax, W. J., Mrabet, N. T., Luiten, R. G. M., Schuurhuizen, P. W., Stanssens, P., & Lasters, I. (1991). Enhancing the thermostability of glucose-isomerase by protein engineering. *Bio/Technology*, *9*(8), 738–742. https://doi.org/10.1038/nbt0891-738
27. Bell, E. L., Finnigan, W., France, S. P., Green, A. P., Hayes, M. A., Hepworth, L. J., Lovelock, S. L., Niikura, H., Osuna, S., Romero, E., Ryan, K. S., Turner, N. J., & Flitsch, S. L. (2021). Biocatalysis. *Nat Rev Methods Primers*, *1*(1), 46. https://doi.org/10.1038/s43586-021-00044-z
28. Ran, N., & Frost, J. W. (2007). Directed evolution of 2-keto-3-deoxy-6-phosphogalactonate aldolase to replace 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase. *J Am Chem Soc*, *129*(19), 6130–6139. https://doi.org/10.1021/ja067330p
29. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., … Newman, J. D. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, *496*(7446), 528–532. https://doi.org/10.1038/nature12051
30. Anastas, P. T., & Warner, J. C. (2000). *Green chemistry : Theory and practice*. Oxford University Press.
31. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., & Hughes, G. J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, *329*(5989), 305–309. https://doi.org/10.1126/science.1188934

Problems

1. Goals