DDT Vol. 8, No. 5 March 2003

Rational design and engineering of therapeutic proteins

Shannon A. Marshall, Greg A. Lazar, Arthur J. Chirino and John R. Desjarlais

An increasing number of engineered protein therapeutics are currently being developed, tested in clinical trials and marketed for use. Many of these proteins arose out of hit-and-miss efforts to discover specific mutations, fusion partners or chemical modifications that confer desired properties. Through these efforts, several useful strategies have emerged for rational optimization of therapeutic candidates. The controlled manipulation of the physical, chemical and biological properties of proteins enabled by structure-based simulation is now being used to refine established rational engineering approaches and to advance new strategies. These methods provide clear, hypothesis-driven routes to solve problems that plague many proteins and to create novel mechanisms of action. We anticipate that rational protein engineering will shape the field of protein therapeutics dramatically by improving existing products and enabling the development of novel therapeutic agents.

Shannon A. Marshall
Greg A. Lazar
Arthur J. Chirino
John R. Desjarlais*
Xencor
111 W. Lemon Ave.
Monrovia
CA 91016, USA
tel: +1 626 305 5900
fax: +1 626 305 0350
*e-mail: jrd@xencor.com

▼ The exquisite specificity of biological therapeutics for their clinical targets has led to their continued development and application as medicines, despite competition from small molecule drugs. Several engineered protein therapeutics are currently being marketed (Table 1), and the annual sales of protein therapeutics are projected to exceed US$59 billion in 2010, which is twice the revenue generated in 2001 (http://www.pharmafile.com/Pharmafocus/Features/feature.asp?fID=281) [1]. For well-validated targets, naturally occurring protein interaction partners constitute preselected 'lead' compounds with high affinity and specificity. However, because natural proteins are not evolved for utilization as drugs, lead optimization is frequently beneficial for development of a protein therapeutic. Modifications can influence the mechanism of action, side effects and efficacy, and satisfy practical constraints such as production costs, intellectual property and dosing frequency.

A variety of strategies have emerged for modulating protein properties, such as efficacy, stability, specificity, immunogenicity and pharmacokinetics (PK). Mechanisms for altering these properties include manipulation of primary structure, incorporation of chemical and post-translation modifications and utilization of fusion partners. The most common route to optimization is site-directed mutagenesis, which is often performed in a brute force or trial-and-error manner. A smaller number of examples exist whereby semirational application of diversity methods, such as phage display, has been used to optimize a therapeutic candidate. Important recent developments are the creation and successful application of rational protein design methods and the determination of an increasing number of high-resolution protein structures.

For the purposes of this review, we define rational protein engineering as the hypothesis-driven manipulation of protein sequence and/or composition. The controlled modification of specific biophysical properties of proteins can potentially impact a variety of therapeutic features (Table 2). An important subset of rational engineering methods consists of approaches that utilize high-resolution, 3D structure information. The most sophisticated of these methods offers an extraordinary level of control over protein sequence and structure, a mechanism to explore sequence combinations that extends far beyond natural diversity, and the ability to couple multiple constraints algorithmically for

 1359-6446/03/$ – see front matter PII: S1359-6446(03)02610-2

Table 1. Engineered protein therapeutics on the market[a]

Name	Family	Company	Indication	Modification	Property
Proleukin® (aldesleukin)	IL-2	Chiron	Cancer	Mutated free cysteine	Decreased aggregation; improved bioavailability
Betaseron® (interferon beta-1b)	IFN-β	Berlex/Chiron	Multiple sclerosis	Mutated free cysteine	Decreased aggregation
Humalog® (insulin lispro)	Insulin	Eli Lilly	Diabetes	Monomer not hexamer	Fast acting
NovoLog® (insulin aspart)	Insulin	Novo Nordisk	Diabetes	Monomer not hexamer	Fast acting
Lantus® (insulin glargine)	Insulin	Aventis	Diabetes	Precipitates in dermis	Sustained release
Enbrel® (etanercept)	TNF receptor	Immunex/ Amgen/Wyeth	Rheumatoid arthritis	Fc fusion	Longer serum half-life; increased avidity
Ontak® (denileukin diftitox)	Diptheria toxin-IL-2	Seragen/Ligand	Cancer	Fusion	Targets cancer cells
PEG-Intron® (peginterferon alfa-2b)	IFN-α	Schering-Plough	Hepatitis	PEGylation	Increased serum half-life; weaker receptor binding
PEGasys® (peginterferon alfa-2a)	IFN-α	Roche	Hepatitis	PEGylation	Increased serum half-life; weaker receptor binding
Neulasta™ (pegfilgrastim)	G-CSF	Amgen	Leukopenia	PEGylation	Increased serum half-life
Oncaspar® (pegaspargase)	Asparaginase	Enzon	Cancer	PEGylation	Decreased immunogenicity; increased serum half-life
Aranesp® (darbepoetin alfa)	Epo	Amgen	Anemia	Additional glycosylation sites	Increased serum half-life; weaker receptor binding
Somavert® (pegvisomant)	Growth hormone	Genentech/ Seragen/ Pharmacia	Acromegaly	PEGylation; binding site mutations	Novel mode of action; increased serum half-life

Chiron (http://www.chiron.com); Berlex (http://berlex.com); Eli Lilly (http://www.lilly.com); Novo Nordisk (http://www.novonordisk.com); Aventis (http://www.aventis.com); Immunex/Amgen (http://www.amgen.com); Wyeth (http://www.wyeth.com); Seragen/Ligand (http://www.ligand.com); Schering-Plough (http://www.sch-plough.com); Roche (http://www.roche.com); Enzon (http://www.enzon.com); Genentech (http://www.genentech.com); Pharmacia (http://www.pharmacia.com).

[a]Abbreviations: G-CSF, granulocyte-colony stimulating factor; IFN-α, interferon α; IL-2, interleukin 2; PEG, polyethylene glycol; TNF, tumor necrosis factor.

simultaneous optimization of several protein properties. Furthermore, proven hypotheses can be reapplied to additional protein systems, thus saving discovery cost and time. Rational methods can be distinguished from those that rely on random sequence perturbations or combinations, such as the class of optimization techniques referred to as directed evolution, although some implementations of these methods have a rational component [2,3].

Physicochemical properties

The physical and chemical properties of protein therapeutics significantly determine their performance during development, manufacturing and clinical use. Many therapeutically interesting proteins are naturally expressed at low concentrations and are degraded rapidly. By contrast, fully developed protein therapeutics require high levels of solubility as well as retention of activity through purification, formulation, storage and administration. Several rational design and engineering strategies, such as those highlighted in Figue 1, have been developed to improve properties such as solubility and stability while maintaining desired biological activity.

Stability

Protein therapeutics are exposed to a variety of stresses that can cause protein unfolding or degradation. Using rational optimization methods, proteins can be re-engineered

Table 2. The biophysical properties of proteins that can be optimized to obtain desired therapeutic outcomes[a]

	Enable discovery	Mechanism of action	Pharmaco-kinetics	Immuno-genicity	Route of administration	Cost of goods	Shelf life	Intellectual property
Stability	x				x	x	x	
Solubility	x		x	x	x	x	x	
Receptor binding affinity and specificity		x	x					
MHC binding affinity			x	x				
Oligomerization state		x	x		x			
Chemical modifications		x	x	x	x		x	x
Posttranslational modifications		x	x	x			x	x
Sequence diversity								x
Conformational state		x						

[a]Abbreviation: MHC, major histocompatibility complex.

such that their structure and activity are substantially more robust with respect to protease exposure, oxidative stress and changes in temperature, pH and solution conditions. One simple stabilization strategy is to replace free cysteines, thereby preventing the formation of unwanted intermolecular or intramolecular disulfide bonds. Cysteine to serine mutations have been introduced successfully into several therapeutic proteins, including granulocyte colony-stimulating factor (G-CSF) and interferon (IFN) β1b, resulting in a longer shelf life [4,5]. Cysteine to serine mutations have also been shown to increase the half-life of human fibroblast growth factor (FGF) [6]. Interestingly, each of the three FGF mutations decreases the thermal stability of the protein, probably because the introduced serines are substantially desolvated and are not positioned to form intramolecular hydrogen bonds. Rational approaches can identify the amino acids that are more precisely compatible with the local structural environment.

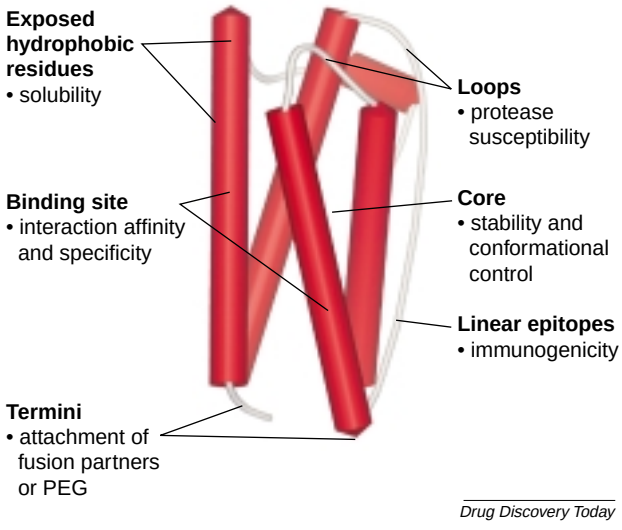

Figure 1. Examples of rational design and engineering strategies. Many design strategies target specific residues or regions of a protein structure for optimization. However, it is important to note that modifications in any region of the protein could potentially affect a wide range of protein properties, emphasizing the importance of rational design methods that can simultaneously consider and optimize multiple parameters. Abbreviation: PEG, polyethylene glycol.

Dramatic improvements in the global stability of a protein can be obtained by optimizing intramolecular interactions. Early examples used rational computational design methods to optimize packing interactions and hydrophobic burial in the protein core [7–9]. Optimizing secondary structure propensity, hydrogen bonds and electrostatic interactions can also improve protein stability substantially [10–12]. More recently, these principles have been applied to the clinically relevant proteins G-CSF and human growth hormone (hGH) by using Protein Design Automation® (PDA™) technology (Box 1). The designed hGH variants are active in cell proliferation studies and are up to 16°C more thermostable than the wild type protein [13]. Optimized G-CSF variants with 10–14 mutations display enhanced thermal stability and five to tenfold increases in shelf life while maintaining the desired biological activity [14].

An additional stabilization strategy is reduction of proteolytic susceptibility. If a specific site in the protein is known to be especially prone to proteolysis, it can be modified so that it no longer matches the substrate specificity of the putative protease. The protease cleavage sites are often

Box 1. Xencor's PDA™ technology: a state of the art rational engineering platform

PDA™ technology, originally developed at Caltech [10,66,67] and further optimized at Xencor [13,14], couples computational design algorithms that generate quality sequence diversity with experimental high-throughput screening to discover proteins with improved properties. The computational component uses atomic level scoring functions, side chain rotamer sampling and advanced optimization methods to capture the relationships between protein sequence, structure and function accurately. Calculations begin with the 3D structure of the protein and a strategy to optimize one or more properties of the protein. PDA™ technology then explores the sequence space comprising all pertinent amino acids (including unnatural amino acids, if desired) at the positions targeted for design. This is accomplished by sampling conformational states of allowed amino acids and scoring them using a parameterized and experimentally validated function that describes the physical and chemical forces governing protein structure. Powerful combinatorial search algorithms are then used to search through the initial sequence space, which can constitute 10^{50} sequences or more, and quickly return a tractable number of sequences that are predicted to satisfy the design criteria. Useful modes of the technology span from combinatorial sequence design to prioritized selection of optimal single site substitutions. PDA™ technology has been applied to numerous systems including important pharmaceutical and industrial proteins and has a demonstrated record of success in protein optimization.

located in flexible loops; therefore, another approach is to introduce mutations that decrease flexibility. Thrombolytics are a class of protein therapeutics for which proteolytic susceptibility is especially important because many clotting factors are both activated and inactivated by specific proteases. For example, an engineered variant of coagulation factor VIIIa with increased resistance to proteolytic inactivation was generated by mutating two arginines required for cleavage by thrombin, factor Xa and activated protein C [15].

Solubility

Protein therapeutics are typically expressed, formulated and administered at high concentrations. Under such conditions many proteins form inclusion bodies during expression or aggregates after formulation. Improving the solubility of a protein can facilitate discovery efforts, whereas enabling soluble prokaryotic expression can reduce production costs dramatically and increase yields. It is far more critical to ensure the solubility of a protein therapeutic once it is administered. Aggregation can cause decreased activity, decreased bioavailability and increased immunogenicity. Several strategies have been applied successfully to reduce protein aggregation and enable soluble expression. Replacement of unpaired cysteine residues can prevent the formation of unwanted intermolecular disulfide bonds, as described above. Post-translational and chemical modifications, which are discussed in a later section, can also help to prevent aggregation. Substituting exposed nonpolar residues with polar residues can enable soluble expression and improve the solubility of the purified protein. This strategy was applied successfully to the A1 domain of cholera toxin, a powerful adjuvant. Of the six variants produced, one retained full biological activity and stability and also displayed a significant improvement in solubility [16]. Altering the net charge and isoelectric point (pI) of a protein can also affect its solubility. For example, a single chain antibody targeting renal cell carcinoma was altered to increase solubility by adding five glutamic acid residues to the C-terminus, thus lowering the pI from 7.5 to 6.1 [17]. Although there are a few examples of rational solubility engineering, the majority of the published successes in solubility optimization have been anecdotal. Until now solubility obstacles have been more or less considered to be formulation problems that can be surmounted with an exhaustive protein chemistry effort. We anticipate that systematic structure-guided optimization efforts will lead to the emergence of well-defined strategies that consistently yield proteins with improved solubility and minimally perturbed structure and function.

Pharmacokinetics

The first generation of protein therapeutics frequently suffered from poor PK. As our understanding of protein clearance processes improves, it becomes possible to rationally modify proteins to tailor their PK profiles. Properly controlling the serum concentration of a therapeutic protein over time can lead to improved efficacy and decreased side effects. In fact, improvements in PK properties can be so vital to the efficacy of a protein drug that they are often made at the expense of specific activity. In addition to eliminating proteolytic susceptibility (see above), several strategies have been developed to alter PK, including polyethlylene glycol (PEG) attachment (PEGylation), glycosylation, fusion to proteins with long serum half-lives, alteration of oligomerization state and modulation of receptor-mediated uptake and

turnover. Knowledge of the dominant route or routes of elimination for a given protein can help significantly in determining which of these strategies will be the most appropriate. For low molecular weight protein therapeutics, kidney filtration dominates, encouraging modifications that increase the effective size. In the case of ligand-receptor systems PK often depends on the relative influence of receptor-mediated clearance versus renal clearance. Affinity and specificity modifications are a central component of many therapeutic optimization strategies, and thus receptor-mediated clearance might play an important role in the efficacy of many proteins, even when it is not considered explicitly.

Fusion proteins

In a straightforward application of molecular size manipulation, proteins covalently fused to themselves often display significantly improved PK profiles [18]. The PK of a therapeutic protein can be increased more dramatically through fusion to a protein that is known to have a long serum half-life, typically albumin or the Fc region of antibodies. Amgen and Wyeth's Enbrel® (etanercept), which is currently marketed for the treatment of rheumatoid arthritis, is a fusion protein consisting of the extracellular domain of p75 tumor necrosis factor receptor (TNFR) and the Fc domain of human IgG. Fc increases the serum half-life of Enbrel®, presumably by both increasing its size and mediating endosomal recycling (see below). Furthermore, because Fc is a dimer, Enbrel®'s affinity for TNF-α is 50- to 1000-fold higher than the affinity of monomeric TNFR [19]. Albumin fusions have been used to generate variants of the anticoagulant proteins hirudin [20] and barbourin [21]. An interesting twist on this approach is to tag proteins with a peptide sequence that specifically binds albumin. Addition of an albumin-binding peptide tag to the antitissue factor D3H44 Fab increases its half-life by approximately 40-fold [22].

Alteration of oligomerization state

The rate of absorption after injection can be affected by the molecular weight and solubility of a protein. An interesting example is provided by comparing wild type insulin, fast-acting insulin variants and sustained-release insulin variants. Native insulin forms a mixture of dimers and hexamers. The fast-acting insulin variants produced by Eli Lilly and Novo Nordisk, Humalog® (insulin lispro) and NovaLog® (insulin aspart) respectively, contain mutations that decrease oligomerization and, therefore, increase the rate of absorption. As a result, patients can administer these fast-acting insulin variants at mealtimes rather than 1 hour before as was required with native insulin. Long-acting insulin variants are used to maintain steady basal insulin levels. For example, the Aventis product Lantus® (insulin glargine) was engineered by increasing the pI to promote precipitation upon subcutaneous injection, thus slowing the rate of absorption [23].

PEGylation

PEG is a highly flexible and soluble polymer that has gained widespread scientific and regulatory acceptance as a chemical modification for therapeutic proteins. PEGylation improves PK predominantly by increasing the effective size of a protein, with most significant effects for proteins smaller than 70 kDa [24,25]. PEGylation can also reduce immunogenicity and aggregation [26]. Although a variety of chemistries exist [27,28] for coupling PEGs of various sizes to proteins, the greatest attachment specificity generally arises from PEGylation at the N-terminus or unpaired cysteines.

Several PEGylated protein therapeutics, such as Schering-Plough's PEG-Intron® (peginterferon alfa-2b) and Roche's PEGasys® (peginterferon alfa-2a), are currently on the market or in late-stage clinical trials. PEGasys® exhibits a 50- to 70-fold increase in serum half-life and substantially reduced variability in serum concentration [29]. A common negative effect of PEGylation, exemplified by both PEGylated IFNs [29,30], is a loss of specific activity. Future studies on these and other proteins should, therefore, focus on minimizing activity loss by optimizing the sites and sizes of PEG attachment rationally.

Glycosylation

Site-specific incorporation of glycosylation sites serves as an additional approach for improving PK. A notable example is Amgen's hyperglycosylated erythropoietin (Epo) variant Aranesp® (darbepoetin alfa), engineered to contain two additional N-linked glycosylation sites. The additional glycosylation increases the serum half-life threefold while reducing *in vitro* binding roughly fourfold [31]. Thus, Aranesp® is another example of how modification can improve *in vivo* efficacy, despite reducing specific activity. Accordingly, future efforts could benefit from using rational methods to identify N-linked or O-linked glycosylation sites that best maintain the structural and functional properties of the protein.

Endocytic trafficking

The PK of many proteins that bind cell-surface receptors can also be affected by endocytic trafficking. Cell-surface receptors and bound ligands are continually internalized by endocytosis. The receptors and ligands can be recycled back to the surface, degraded in lysozomes or transported

across cells (e.g. from the apical membrane to the basolateral membrane). The fate of the ligand is often determined by the extent of association with the receptor within the endosome, although the relationship between processing and association is highly system dependent. The pH in endosomal compartments is lower than that of serum. Because protein–protein interactions are typically pH-dependent, many ligands are freed from their receptors as they proceed through the endosomal pathway. In some protein families released ligand is recycled, whereas ligand that remains bound is targeted for degradation. For example, the more tightly epidermal growth factor (EGF) family ligands bind EGF receptor at pH 6, then the lower the fraction of ligand that is recycled [32]. Recently Lauffenberger and colleagues have taken advantage of the pH dependence of endosome-mediated G-CSF turnover to rationally engineer variants with improved PK. Several residues involved in receptor binding were mutated to histidine, which is neutrally charged in serum but has a net positive charge in the acidic environment of endocytic vesicles. The variants were predicted to bind receptor normally at the cell surface but to release more effectively than the wild type after endocytosis. Two mutations were shown to have increased half-life and potency compared with wild type G-CSF [33]. This elegant example illustrates the ability of rational engineering methods to use accumulated biological knowledge to generate improved therapeutics. Another important example, discussed below in more detail, is the pH-dependent recycling of immunoglobulin Fc domains. In this case the effect is opposite: the pH drop purges antigen from the variable region while enhancing Fc binding to its receptor, thus enabling the antibody and its receptor to be recycled to the serum.

Affinity, specificity and conformational control

Rational design can be used to modify the affinity and specificity of interactions between a therapeutic protein and other biomolecules. In some cases, increasing the binding affinity for a target protein can produce an increase in biological activity. In other cases, it is possible to reduce undesired biological activities by decreasing the affinity for nontarget molecules. An example of affinity enhancement is the generation of superagonist variants of human thyrotropin (hTSH) by altering the net charge of the protein. The hTSH receptor has a net negative charge, and mutations that introduce positively charged residues or replace negatively charged residues in the peripheral loops of hTSH increase activity. The best variants show a 50,000-fold increase in receptor binding affinity and 1000-fold increase in *in vivo* activity [34,35].

The power of rational design is most impressive when it is used to generate novel mechanisms of action. For example, 4-helix bundle cytokines, including vascular endothelial growth factor (VEGF), hGH and interleukin-6 (IL-6), have been engineered to function as receptor antagonists rather than agonists. Most members of the 4-helix bundle cytokine family must form multiple protein–protein interactions at the cell surface to trigger signaling. VEGF, for example, forms homodimers that bind to two VEGF receptors, whereas IL-6 binds to a low-affinity IL-6 co-receptor and gp130. Antagonistic VEGF variants were designed as heterodimers, which contain one functional binding site per dimer [36]. An IL-6 superantagonist was generated by selecting mutations that disrupt binding to gp130 and incorporating mutations that result in increased affinity for the IL-6 co-receptor [37]. An especially interesting example of a designed cytokine antagonist is Genentech/Pharmacia's Somavert® (pegvisomant), a hGH variant that has recently successfully completed clinical trials for treatment of acromegaly. hGH contains two distinct receptor binding sites and dimerizes its receptor upon binding. Somavert® contains a point mutation at the second receptor binding site that blocks receptor dimerization [38] and eight additional mutations, identified by phage display, that increase the receptor-binding affinity of the first site [39].

Many proteins undergo conformational changes that are central to their function. In such cases, rational design methods can drive conformational equilibria towards the therapeutically desirable state. A notable example is the design of constitutively active and inactive integrin I domain variants. Integrin I domains can populate two dominant conformations: an 'open' conformation, which can bind intracellular adhesion molecule-1 (ICAM-1), and a 'closed' conformation, which has very low affinity for ICAM-1. The native protein rests in the closed conformation and converts to the open conformation during signaling. Springer and coworkers used two distinct strategies to generate conformationally locked integrin I domain variants. One approach introduced pairs of cysteines that form disulfide bonds compatible with either the closed or open conformation [40,41]. In the second approach, mutations were designed in the core of the domain that were computationally selected to stabilize the open conformation and disallow the closed state [42].

Immunogenicity

The potential for protein therapeutics to produce harmful immune responses is a significant barrier to the development and acceptance of protein drugs. The immune response is typically most severe for nonhuman proteins. For example, antibodies against streptokinase, a bacterially derived

Table 3. Engineered antibodies on the market[a]

Name	Company	Target	Indication	Type
Orthoclone OKT3® (muromonab-CD3)	Ortho Biotech/ Johnson & Johnson	CD3	Transplant rejection	Murine
ReoPro® (abciximab)	Centocor/Lilly	GPIIb/IIIa	Restenosis	Chimeric
Rituxan® (rituximab)	IDEC/Genentech	CD20	B-cell non-Hodgkins lymphoma	Chimeric
Simulect® (basiliximab)	Novartis	IL-2R	Transplant rejection	Chimeric
Remicade® (infliximab)	Centocor	TNF-α	Crohn's disease, rheumatoid arthritis	Chimeric
Zevalin® (ibritumomab tiuxetan)	IDEC/Schering AG	CD20	B-cell non-Hodgkins lymphoma	Chimeric
Zenapax® (daclizumab)	PDL/Roche	IL-2R	Transplant rejection	Humanized
Synagis® (palivizumab)	MedImmune	RSV F protein	Respiratory syncitial virus	Humanized
Herceptin® (trastuzumab)	Genentech	HER2/neu	Breast cancer	Humanized
Mylotarg® (gemtuzumab ozogamicin)	Celltech/Wyeth	CD33	Acute myeloid leukemia	Humanized
Campath® (alemtuzumab)	Millenium/ILEX	CD52	B-cell chronic lymphocytic leukemia	Humanized

Ortho Biotech (http://www.orthobiotech.com); Johnson & Johnson (http://www.jnj.com); Centocor (http://www.centocor.com); Eli Lilly (http://www.lilly.com); IDEC (http://www.idec.com); Genentech (http://www.genentech.com); Novartis (http://www.novartis.com); Schering Ag (http://www.schering.de/eng/); PDL (http://www.pdl.com); Roche (http://www.roche.com); MedImmune (http://www.medimmune.com); Celltech (http://www.celltechgroup.com); Wyeth (http://www.wyeth.com); Millenium (http://www.mlnm.com); ILEX (http://www.ilexonc.com).

[a]Abbreviations: GPIIb/IIIa, platelet glycoprotein IIb/IIIa; IL-2R , interleukin 2 receptor; RSV F, respiratory syncitial virus F; TNF-α, tumor necrosis factor α.

antithrombolytic used to treat myocardial infarction, not only neutralize the protein and reduce its efficacy, but also can elicit severe allergic reactions that effectively limit streptokinase therapy to one-time use. Yet even therapeutics based on human proteins can cause immune responses depending on the mode of administration (including dosage, frequency and route) and the solubility and stability of the formulated protein. Neutralizing antibodies have been observed against a variety of human proteins including insulin, factor VIII, IFNs, Epo and megakaryocyte growth and differentiation factor (MGDF). In some cases, for example with the multiple sclerosis drug IFN-β, efficacy is severely hindered due to neutralizing antibodies [43]. Devastating problems can result when elicited antibodies crossreact with endogenous protein. For example, clinical trials of MGDF were halted when crossreactive neutralizing antibodies to endogenous thrombopoietin caused reduced platelet counts (thrombocytopenia) in a small number of otherwise healthy volunteers. As another example, Johnson and Johnson's European formulation of Epo, Eprex® (epoetin alfa) has caused pure red blood cell aplasia in several patients owing to the formation of crossreactive neutralizing antibodies.

The application of rational engineering to immunogenicity has been aimed mostly at increasing the antigenicity of proteins for use in vaccines. Immune reduction of proteins as a whole is not as straightforward, and relatively few examples exist of rationally reducing or eliminating the immunogenicity of protein therapeutics. The only real success for immunogenicity reduction has been the humanization of murine antibodies, made possible by the high regularity of antibody sequence and structure and the ability to use proximity to human sequence as a metric for immunogenicity. In some cases, PEGylation has reduced the fraction of patients who raise neutralizing antibodies, possibly by sterically blocking access to epitopes [44]. Rational design methods that improve the solution properties of a protein therapeutic might also reduce immunogenicity because aggregates are generally more immunogenic than soluble proteins.

A more general approach to de-immunization involves mutagenesis of epitopes in the protein sequence and structure that are most responsible for stimulating the immune system. Some success has been achieved by randomly replacing surface residues, thus generating sequences with lower affinity for panels of known neutralizing antibodies [45,46]. An alternate approach is to disrupt T-cell activation by mutating peptides that bind class II major histocompatibility complex (MHC) alleles. Removal of MHC-binding epitopes offers a much more tractable approach to de-immunization than the removal of antibody epitopes because the diversity of MHC molecules comprises only $1–2x10^3$ alleles, whereas the antibody repertoire is estimated to be approximately 10^8. A current challenge for rational design methods is to identify sequence variants that eliminate potential MHC-binding epitopes while maintaining protein

structure and function. We anticipate that a general solution to the problem of protein immunogenicity will improve the safety and efficacy of protein therapeutics substantially and will enable new classes of nonhuman and *de novo* proteins to enter the clinic.

Antibodies

Some of the most visible and successful applications of rational engineering methods to biotherapeutics have occurred in the field of antibodies. Monoclonal antibodies are widely used as treatments for a variety of conditions from arthritis to cancer. There are currently 11 antibody products on the market, as shown in Table 3, and well over 100 in development. Despite such widespread acceptance and promise, there is still a need for structural and functional antibody optimization. Current antibody engineering efforts target both the variable and Fc regions of the molecule.

Antibody variable domains suffer from stability and solubility issues similar to all proteins, as discussed previously. However, because antibodies share a common structural scaffold, rational engineering studies have been able to dissect some of the sequence and structural determinants of variable region solubility and stability [47]. Notable developments include the structure-based design of more finely tuned complementarity determining region grafts and libraries [48,49], the use of phage-based selection methods for humanization [50,51] or fully human antibody generation [52] and the application of computational methods to increase the association rate of antibody/antigen formation at predicted 'ON-Rate AMPlification Sites' (Marvin and Lowman, pers. commun.). Furthermore, owing to the modular nature of immunoglobulin domains, variable domain architectures, such as diabodies, triabodies and bispecific diabodies, are being engineered to better serve specific therapeutic applications [53,54]. For more detail on variable region engineering the reader is referred to an excellent review by Maynard and Georgiou [53].

The Fc region of an antibody mediates interactions with several receptors, thus allowing antibodies to recruit the immune system and possess an extended serum half-life [55,56]. Significant effort has gone into engineering Fc for enhanced functional properties. Most exciting are recent results indicating that tighter binding by Fc to certain Fc gamma receptors, obtained by mutagenesis [57] or expression of carbohydrate isoforms [58,59], can result in enhanced effector function, potentially enabling the engineering of more potent antitumor antibodies. Additionally, some success has been achieved in modulating antibody PK by generating Fc variants with altered affinity for the neonatal receptor FcRn [60,61]. The bottleneck for Fc engineering is production. Because of the requirement for glycosylation, Fc and full-length antibodies must be produced in mammalian systems, precluding screening of large numbers of variants. Engineering a system with such high therapeutic potential yet limited screening capacity will be an exciting challenge for rational protein design.

State of the art rational engineering

The numerous examples discussed in this review illustrate both the demand for and power of rational engineering methods to improve the efficacy of biotherapeutics. There is currently an opportunity to replace the typical hit-and-miss approach to protein optimization with quantitative and systematic engineering strategies using computational

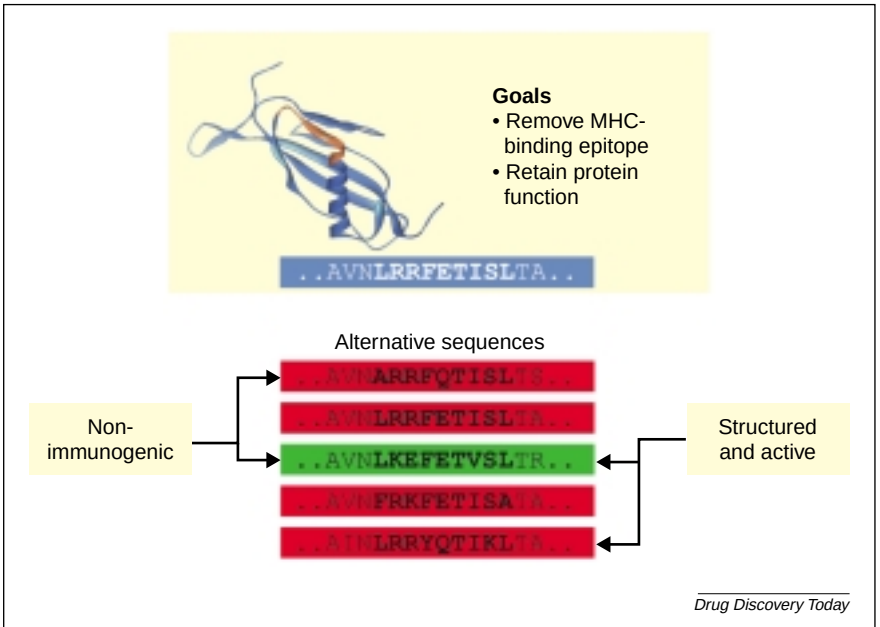

Figure 2. Multiparameter optimization of proteins. The modification of linear sequence epitopes is useful for modulating immunogenicity, altering proteolytic stability, introducing chemical modification sites and other strategies. However, these changes have a high likelihood of disrupting the integrity of the structure and function of the protein. It can be difficult to experimentally select for multiple properties simultaneously, such as nonimmunogenic and active variants. Rational design approaches are uniquely suited to the identification of protein sequences that satisfy multiple constraints. Abbreviation: MHC, major histocompatibility complex.

protein design methods [62–65]. Xencor's PDA™ technology is an example of these new methods (Box 1).

The full potential of computational design algorithms is realized when they are followed by high-throughput experimental screening efforts to single out superior members of a protein library. Computationally generated libraries are significantly enriched in stable, properly folded sequences relative to randomly generated libraries. In effect, structure-based sequence sampling methods yield an increased hit-rate, thereby decreasing the number of variants that must be screened. This feature is often critical to success because screens for therapeutic proteins, such as cell-based or *in vivo* assays, are often extremely low throughput. Given a high quality library, experimental screening methods can identify the sequence or sequences with the best characteristics quickly.

Computational design algorithms have tremendous potential for addressing conflicting constraints on a protein's sequence and structure, a common challenge in protein optimization efforts. As illustrated in Figure 2, many strategies (e.g. introduction of chemical or post-translational modification sites, removal of proteolysis sites and removal of MHC epitopes) require modifications to local primary structure. In most cases, however, the effect of these changes on the tertiary structure and functional integrity of the protein must also be considered. In other cases, one seeks primary structure alterations that disrupt one interaction while preserving a multiplicity of other interactions. Unfortunately, the number of acceptable sequence solutions narrows dramatically as the number of constraints is increased. One costly solution to this general problem is to develop assays that assess compatibility with each constraint separately. Alternatively, computational algorithms can simultaneously consider most or all of the constraints in the context of the whole protein. This approach also affords the opportunity to discover compensatory mutations elsewhere in the protein to accommodate changes made at the primary optimization site.

Conclusions

To convert a typical endogenous protein to a successful therapeutic it is often necessary to optimize several parameters, such as stability, solubility, PK and immunogenicity, while preserving or even enhancing function. Many strategies have already emerged for perturbing these parameters. We anticipate that the continued development and application of rational protein design technology will enable significant improvements in the efficacy and safety of existing protein therapeutics, as well as allow the generation of entirely novel classes of proteins and modes of action.

Acknowledgements

We would like to thank Bassil Dahiyat and David Szymkowski for helpful conversations and editorial advice.

References

1 Pharmafocus (2002) *Biotechnology 2006: Shaping the Future.* Available at: http://www.pharmafile.com/Pharmafocus/Features/feature.asp?fID=281
2 Voigt, C.A. *et al.* (2002) Protein building blocks preserved by recombination. *Nat. Struct. Biol.* 9, 553–558
3 Voigt, C.A. *et al.* (2001) Computationally focusing the directed evolution of proteins. *J. Cell. Biochem. Suppl.* 37, 58–63
4 Lin, L. (1998) Betaseron. *Dev. Biol. Stand.* 96, 97–104
5 Arakawa, T. *et al.* (1993) Cysteine 17 of recombinant human granulocyte colony stimulating factor is partially solvent-exposed. *J. Protein Chem.* 12, 525–531
6 Culajay, J.F. *et al.* (2000) Thermodynamic characterization of mutants of human fibroblast growth factor 1 with an increased physiological half-life. *Biochemistry* 39, 7153–7158
7 Desjarlais, J.R. and Handel, T.M. (1995) *De novo* design of the hydrophobic cores of proteins. *Protein Sci.* 4, 2006–2018
8 Hellinga, H.W. and Richards, F.M. (1994) Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5803–5807
9 Hurley, J.H. *et al.* (1992) Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. *J. Mol. Biol.* 244, 1143–1159
10 Dahiyat, B.I. *et al.* (1997) Automated design of the surface positions of protein helices. *Protein Sci.* 6, 1333–1337
11 Malakauskas, S.M. and Mayo, S.L. (1998) Design, structure, and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* 5, 470–475
12 Marshall, S.A. *et al.* (2002) Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* 316, 189–199
13 Filikov, A.V. *et al.* (2002) Computational stabilization of human growth hormone. *Protein Sci.* 11, 1452–1461
14 Luo, P. *et al.* (2002) Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening. *Protein Sci.* 11, 1218–1226
15 Pipe, S.W. and Kaufman, R.J. (1997) Characterization of a genetically engineered inactivation-resistant coagulation factor VIIIa. *Proc. Natl. Acad. Sci. U. S. A.* 94, 11851–11856
16 Agren, L. *et al.* (1999) Hydrophobicity engineering of cholera toxin A1 subunit in the strong adjuvant fusion protein CTA1-DD. *Protein Eng.* 12, 173–178
17 Tan, P.H. *et al.* (1998) Engineering the isoelectric point of a renal cell carcinoma targeting antibody greatly enhances scFv solubility. *Immunotechnology* 4, 107–114
18 Sytkowski, A.J. *et al.* (1999) An erythropoietin fusion protein comprised of identical repeating domains exhibits enhanced biological properties. *J. Biol. Chem.* 274, 24773–24778
19 Goldenberg, M.M. (1999) Etanercept, a novel drug for the treatment of patients with severe, active rheumatoid arthritis. *Clin. Ther.* 21, 75–87
20 Syed, S. *et al.* (1997) Potent antithrobmin activity and delayed clearance from the circulation characterize recombinant hirudin genetically fused to albumin. *Blood* 89, 3243–3252
21 Marques, J.A. *et al.* (2001) A barbourin-albumin fusion protein that is slowly cleared *in vivo* retains the ability to inhibit platelet aggregation *in vitro*. *Thromb. Haemost.* 86, 902–908
22 Dennis, M.S. *et al.* (2002) Albumin-binding as a general strategy of improving the pharmacokinetics of proteins. *J. Biol. Chem.* 277, 35035–35043
23 Vajo, Z. and Duckworth, W.C. (2000) Genetically engineered insulin analogs: diabetes in the new millenium. *Pharmacol. Rev.* 52, 1–9
24 Knauf, M.J. *et al.* (1988) Relationship of effective molecular size to systemic clearance in rats of recombinant interleukin-2 chemically modified with water-soluble polymers. *J. Biol. Chem.* 263, 15064–15070

25 Clark, R. *et al.* (1996) Long-acting growth hormones produced by conjugation with polyethylene glycol. *J. Biol. Chem.* 271, 21969–21977
26 Harris, J.M. *et al.* (2001) Pegylation: a novel process for modifying pharmacokinetics. *Clin. Pharmacokinet.* 40, 539–551
27 Roberts, M.J. *et al.* (2002) Chemistry for peptide and protein PEGylation. *Adv. Drug Deliv. Rev.* 54, 459–476
28 Kinstler, O. *et al.* (2002) Mono-N-terminal poly(ethylene glycol)-protein conjugates. *Adv. Drug Deliv. Rev.* 54, 447–485
29 Bailon, P. *et al.* (2001) Rational design of a potent, long-lasting form of interferon: a 40-kDa-branched polyethylene glycol-conjugated interferon alpha-2a for the treatment of hepatitis C. *Bioconjug. Chem.* 12, 195–202
30 Wang, Y.S. *et al.* (2002) Structural and biological characterization of pegylated recombinant interferon alpha-2b and its therapeutic implications. *Adv. Drug Deliv. Rev.* 54, 547–570
31 MacDougall, I.C. *et al.* (1999) Pharmacokinetics of novel erythropoiesis stimulating protein compared with epoetin alfa in dialysis patients. *J. Am. Soc. Nephrol.* 10, 2392–2395
32 French, A.R. *et al.* (1995) Intracellular traffiking of epidermal growth factor family ligands is directly influenced by the pH sensitivity of the receptor/ligand interaction. *J. Biol. Chem.* 270, 4334–4340
33 Sarkar, C.A. *et al.* (2002) Rational cytokine design for increased lifetime and enhanced potency using pH activated histidine-switching. *Nat. Biotechnol.* 20, 908–913
34 Weintraub, B.D. and Szkudlinski, M.W. (1999) Development and *in vitro* characterization of human recombinant thyrotropin. *Thyroid* 9, 447–450
35 Grossmann, M. *et al.* (1998) A rational design strategy for protein hormone superagonists. *Nat. Biotechnol.* 16, 871–875
36 Siemeister, G. *et al.* (1998) An antagonistic vascular endothelial grown factor (VEGF) variant inhibits VEGF-stimulated receptor autophosphorylation and proliferation of human endothelial cells. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4625–4669
37 Savino, R. *et al.* (1994) Rational design of a receptor super-antagonist of human interleukin-6. *EMBO J.* 13, 5863–5870
38 Fuh, C. *et al.* (1992) Rational design of potent antagonists to the human growth hormone receptor. *Science* 256, 1677–1680
39 Olson, K. *et al.* (1998) *Human Growth Hormone Variants*. US patent number 5849535
40 Lu, C. *et al.* (2001) An isolated, surface expressed I domain of the integrin alphaL beta2 is sufficient for strong adhesive function when locked in the open conformation with a disulfide bond. *Proc. Natl. Acad. Sci. U. S. A.* 98, 2387–2392
41 Shimaoka, M. *et al.* (2001) Reversibly locking a protein fold in an active conformation with a disulfide bond: integrin alphaL I domains with high affinity and antagonist activity *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6009–6014
42 Shimaoka, M. *et al.* (2000) Computational design of an integrin I domain stabilized in the open high affinity conformation. *Nat. Struct. Biol.* 7, 674–678
43 Antonelli, G. and Dianzani, F. (1999) Development of antibodies to interferon beta in patients: technical and biological aspects. *Eur. Cytokine Netw.* 10, 413–422
44 He, X.H. *et al.* (1999) Reducing the immunogenicity and improving the *in vivo* activity of trichosanthin by site-directed pegylation. *Life Sci.* 65, 355–368
45 Meyer, D.L. *et al.* (2001) Reduced antibody response to streptavidin through site-directed mutagenesis. *Protein Sci.* 10, 491–503
46 Laroche, Y. *et al.* (2000) Recombinant staphylokinase variants with reduced antigenicity due to elimination of B-lymphocyte epitopes. *Blood* 96, 1425–1432
47 Worn, A. and Pluckthun, A. (2001) Stability engineering of antibody single chain Fv fragments. *J. Mol. Biol.* 305, 989–1010
48 Tamura, M. *et al.* (2000) Structural correlates of an anticarcinoma antibody: identification of specificity-determining residues (SDRs) and development of a minimally immunogenic antibody variant by retention of SDRs only. *J. Immunol.* 164, 1432–1441
49 Chen, Y. *et al.* (1999) Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen. *J. Mol. Biol.* 293, 865–881
50 Rader, C. *et al.* (1998) A phage display approach for rapid antibody humanization: designed combinatorial V gene libraries. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8910–8915
51 Baca, M. *et al.* (1997) Antibody humanization using monovalent phage display. *J. Biol. Chem.* 272, 10678–10684
52 Griffiths, A.D. and Duncan, A.R. (1998) Strategies for selection of antibodies by phage display. *Curr. Opin. Biotechnol.* 9, 102–108
53 Maynard, J. and Georgiou, G. (2000) Antibody engineering. *Annu. Rev. Biomed. Eng.* 2, 339–376
54 Pluckthun, A. and Pack, P. (1997) New protein engineering approaches to multivalent and bispecific antibody fragments. *Immunotechnology* 3, 83–105
55 Ravetch, J.V. and Bolland, S. (2001) IgG Fc receptors. *Annu. Rev. Immunol.* 19, 275–290
56 Raghavan, M. and Bjorkman, P.J. (1996) Fc receptors and their interactions with immunoglobulins. *Annu. Rev. Cell Dev. Biol.* 12, 181–220
57 Shields, R.L. *et al.* (2001) High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R. *J. Biol. Chem.* 276, 6591–6604
58 Shields, R.L. *et al.* (2002) Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* 277, 26733–26740
59 Umana, P. *et al.* (1999) Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. *Nat. Biotechnol.* 17, 176–180
60 Ghetie, V. *et al.* (1997) Increasing the serum persistence of an IgG fragment by random mutagenesis. *Nat. Biotechnol.* 15, 637–640
61 Kim, J.K. *et al.* (1999) Mapping the site on human IgG for binding of the MHC class I-related receptor, FcRn. *Eur. J. Immunol.* 29, 2819–2825
62 Street, A.G. and Mayo, S.L. (1999) Computational protein design. *Structure Fold Des.* 7, R105–109
63 Mendes, J. *et al.* (2002) Energy estimation in protein design. *Curr. Opin. Struct. Biol.* 12, 441–446
64 Pokala, N. and Handel, T.M. (2001) Review: Protein design- Where we were, where we are, where we're going. *J. Struct. Biol.* 134, 269–281
65 Kraemer-Pecore, C.M. *et al.* (2001) Computational protein design. *Curr. Opin. Chem. Biol.* 5, 690–695
66 Dahiyat, B.I. and Mayo, S.L. (1996) Protein design automation. *Protein Sci.* 5, 895–903
67 Dahiyat, B.I. and Mayo, S.L. (1997) *De novo* protein design: fully automated sequence selection. *Science* 278, 82–87