

Protein therapeutics

– new ladders up the fruit tree

Therapeutic proteins are often regarded as too technologically demanding, too costly, too restricted in routes or duration of administration or – in an era of massive opportunity provided by high throughput screening, combinatorial chemistry and genomics – simply passé. This article explores how fundamental advances (such as definition of the human proteome and protein combinatorics) as well as incremental steps forward (such as improvements in production methods) are rapidly expanding therapeutic uses for the biology's most important molecules.

Imagine you are playing a very frustrating game in a very large dark room. You don't even know exactly how big this room is but you do know that scattered all over its floor are objects made of linked shaped units like children's building blocks. You are permitted to grope around and pick up one object at a time, illuminating it with a small and unreliable flashlight which occasionally picks up other similar objects nearby. Before the light goes out, you may move the object to the lit corridor outside where you hire a lawyer to guard it before you return to the darkness. From time to time you hear muffled cries of exultation as one group of lawyers realises that another will pay something for what is in their pile. Then, not instantaneously but quite quickly, the lights come on in the great room. The ensuing scene is reminiscent of the competition prizes once given by toy shops where children were given so many minutes to load into a shopping cart as many toys as they could lay hands on in the store. Small armies of fellow investigators rush about trying to throw whole loads of toys into the protected corridor. Most fall short and as the scene is illuminated, you realise that you (and everyone else) now has a full view of what lies on the floor and more importantly, what all the building blocks are. It is quite a sight: not only can you count the building blocks but there more of some than you ever would have guessed, less of others, types you never suspected and combinations you never would have dreamt of. In that

distribution, you are for the first time seeing the real rule book of the game.

This, very roughly, is the situation of the protein engineer in the post-genomic era.

The market message

The phrase "we are devoting an increasing proportion of our R&D resources to discovery and development of low molecular weight agents" probably sounds familiar to most of the readership of *Drug Discovery World*. When it emanates from the larger and established US biotechnology companies who owe their existence to successful development of protein therapeutics, it is usually interpreted as meaning: **a)** that they are seeking to appear more 'mainstream pharma-like' (a questionable aspiration) and **b)** that they consider that the easy pickings (the 'low-hanging fruit') have been harvested with the result that market growth potential for protein agents is limited.

The latter point has become something of a mantra in pharmaceutical management circles in recent years but it is based on a serious misconception. In 1997 the annual growth in the non-vaccine therapeutic protein market was estimated at around 6% – well down from the giddy heights of 40-60% pa in the early 1990s and some projections suggested negative growth for the 1998-2000 period. In fact, the therapeutic protein market has achieved year-on-year growth in the 20-30% range over this period¹. Annual

By Dr Richard AG
Smith

Therapeutics

Table 1

Selected types of protein domain found in the human genome and their combination

Domain Type	Number Of Genes Encoding This Domain In Human Genome	Number Of Individual Domains In Human Genome	Number Of Domains With Assignable Function (1)	Intra-Set Linear Combinations Of 3-Different Functional Domains (2)
Fibronectin type III	106	545	14	2,184
TGF- α	27	28	11	990
Kringle	15	24	7	210
IL-1	7	7	3	6
Anaphylotoxin	6	14	4	24
Ig	381	930	43	74,046
Zinc fingers (all types)	725	4,676	113	1,404,816

(1) Hypothetical

(2) Based on hypothetical numbers of functional domains; assumes that the order of domain assembly is functionally significant

sales have increased from about \$12 billion in 1997 to ~\$20 billion in 1999¹. It is true that growth is erratic and that haematopoietic factors still account for 20-25% of the market but the overall performance of this sector of the pharmaceutical market compares rather favourably with the conventional chemical agent market. Furthermore, these numbers do not include the protein-based vaccine market which has also shown very healthy growth. It is now possible to identify specific technological drivers for further growth in both the number and the value of protein therapeutics.

Genomics, proteomics and domain combinatorics

The analogy of the toy building blocks (domains) can be applied to the majority of human gene products, the main exceptions being arguably some extended structural proteins such as the collagens. The fact that the human genome contains 'only' around 30,000 genes^{2,3} still leaves the possibility of a much larger number of mRNA transcripts (>85,000 according to some estimates⁴) and when post-translational modification is taken into account, the probability of a very much larger human proteome. If the independently folded protein domain is taken as the basic unit of functionality in the proteome, then it is worth considering how many human protein domains there are and in how many ways they may be combined.

Annotation of the human genome sequence has permitted estimates of the number of domains identifiable by sequence homology^{2,3}. These estimates should be treated with caution because their definitions of domain tend to ignore similarities in three-dimensional structure (eg among many of the interleukins) that are not immediately evident from sequence analysis. However, they do provide at least an outline of the scope for the combinatorial creation of novel functions by engineering of human protein domains. **Table 1** illustrates this in a simple way by giving examples of the number of human genes containing a particular domain type, the total number of such domains identified and a (hypothetical) number of different functions that might be identified in each domain set. Assuming that sequential domain orientation is potentially significant in creating a novel function (ie that NH₂-(A)-(B)-(C)-CO₂H is not the same as NH₂-(C)-(B)-(A)-CO₂H), an estimate is given of the possible number of same-domain constructs of a given size. It is immediately obvious that even with simple two- or three-domain constructs, there are a large number of possibilities. The estimate of the number of different domains involved in adhesion, for example, is around 2,400² and the possible ways of combining these in mere two-domain constructs approaches six million. Therefore as domain-focused functional genomics advances through mapping of the 'interactome', the scope

for rational design of multi-functional protein therapeutics will increase rapidly even before site-directed mutagenesis and pharmacological design is brought into the equation.

Enter the biology: combinatorics and further tinkering

Suppose you want to target a pro-inflammatory agent to a tumour characterised by expression of a marker protein which screening has established binds four of the 24 available human kringle domains. You need a minimum of two kringles to get a suitable level of affinity for this marker so that a chemotactic gradient can be established near the tumour site at sensible therapeutic concentrations. Constructs with more than two kringles are found not to express well in your chosen system (see below). The biologically plausible pro-inflammatory options amount to two of the 14 known anaphylotoxin domains, two cytokines and three chemokines, a total of seven and you are limited by expression considerations to having only one of these per construct. There are 16 possible ways of assembling two kringles if pairs of identical kringles are admissible. The numbers of possible two-kringle, one-inflammatory mediator constructs is 224. This is a significant number of proteins to have to make at scale and evaluate biologically but it is well within the scope of, for example, a screen based on transient expression of GPI-anchored constructs on cells whose interactions with the tumour line is being measured. Functional improvements may be increased by rational site-directed or scanning mutagenesis to optimise binding or activity and therapeutic ratio may be improved by creating a pro-drug form of the pro-inflammatory domain which is activated by a tumour-associated protease. Finally, the whole agent may be given a long plasma half-life by modification with polyethylene glycol, be targeted to tumour cell surfaces by a ligand address and have linker regions designed to minimise immunogenicity. The kind of structure you could finish up with is illustrated in **Figure 1**. Examples of the recent use of these techniques individually can be found with plasminogen activators^{5,6}, TNF-regulators⁷ and complement inhibitors^{8,9}.

Novel chemistry for post-translational modification of proteins and total synthesis of proteins is also developing and this provides a further set of options for construction of new types of targeted and poly-functional molecules (see reference 10 for a recent example of a total synthesis using native chemical ligation).

Table 2

Characteristics of the main protein expression systems

	Scaleability	Yield/unit vol/ Biomass	DSP Issues	Complex Proteins?	Validated for Pharmaceuticals?	Environment & Ethics issues	Cost of Goods (1)
E.coli intracellular	++++	+++(+)	Refolding	In some cases	Yes	No	++++
E.coli secreted	+++	+(+)	Minor	"	Coming	No	+++
Other bacterial	+++	+(+)	Proteases+	"	Limited	No	+++
Yeasts	+++	+(+)	Extraction proteases	Often	Yes	No	++
Filamentous fungi	++++	+	Extraction	Often	Not yet	No	?
Baculovirus/ insect cells	++	++	Immunogenic impurities	Usually	Close	No	++
Mammalian cells (in culture)	+++	++	Viruses, oncogenes	Yes	Yes	No	+
Mammals (milk)	+++(+)	++(+)	Lipids, viruses	Yes	Close	Probable	+++
Birds (eggs)	+++	++	"	Usually	Not yet	Possible	++?
Fish	++(+)	?	Biomass extraction	In principle	No	Possible	?
Plants	+++(+)	+	"	Often	Close	Probable	+++?

++++ Best

+ Worst

(1) Per course of treatment

Production and economics

The general perception that protein therapeutics are costly to manufacture and suffer from cost-of-goods limitations still has a ring of truth but this is an area where steady incremental progress is being made in improving and scaling up established technology and in increasing expression system options. A full review of protein production is far beyond the scope of this article but Table 2 provides a summary of most of the production systems in current use. *E.coli* remains the economic gold standard host cell, largely because intracellular expression can give primary production yields in the multi-gram/L range and fermentation can be scaled to the levels required for the multi-kg batches needed for market entry. However, many proteins cannot be refolded from inclusion bodies and others may require post-translational modification. For these, a large number of options now exist ranging from the tried and tested chinese hamster ovary (CHO) cell to transgenic animals. When designing protein therapeutics, the relationship between total course-of-treatment dosage, expression system and potency needs always to be kept in mind.

If the proposed treatment in man looks like using more than ~1gram/year and the construct needs to be expressed in CHO cells because of its complexity, then a mutagenesis strategy to increase

potency is worth considering at an early stage. Although expression science remains a somewhat empirical undertaking, selection of domains which are individually expressible in *E.coli* and linked so that the domains can refold autonomously is a good starting point for the kinds of construct illustrated in Figure 1.

Delivery

There are good biological reasons why the gut and the skin resist the transport of intact folded proteins across their barriers and into the systemic circulation. Real progress in achieving generally applicable techniques for delivery by these routes remains painfully slow but perceptible advances are being made elsewhere. Again, potency is all: it is much easier to sell a weekly blast of 10µg of a therapeutic going through the dermal layer on supersonic gold particles than it is to sell 100mg by intramuscular injection at a similar frequency. In other words, one of the often cited objections to protein therapeutics – that they have to be given parenterally – is beginning to yield to particle delivery and inhalation approaches. This is particularly important for vaccines but given sufficiently potent combinatorial agents, may be also be a key growth point for therapeutic proteins.

References

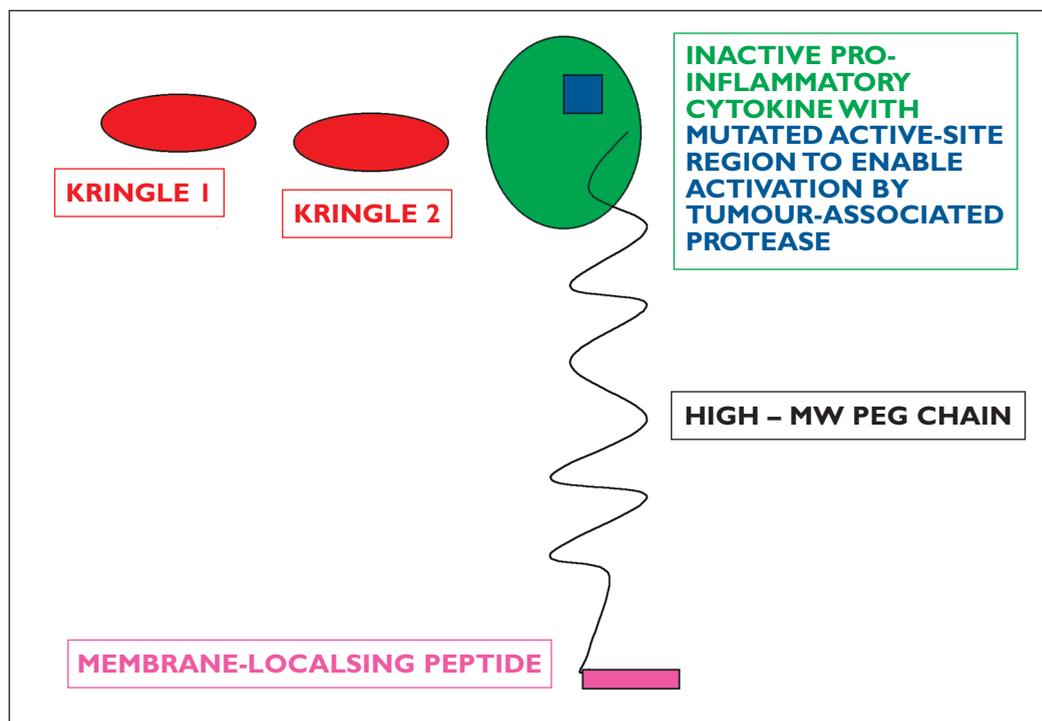
- 1 Datamonitor reports on therapeutic proteins 1997-1999, www.datamonitor.com.
- 2 Venter, JC et al. The sequence of the human genome. *Science*. 2001, 291, 1304-1351.
- 3 Lander, ES et al. Initial sequencing and analysis of the human genome. *Nature*. 2001, 409, 860-921.
- 4 Claverie, J-M. What if there are only 30,000 human genes? *Science*. 2001, 291, 1255-1257.
- 5 Higgins, DL, Bennet, WF. Tissue plasminogen activator: the biochemistry and pharmacology of variants produced by mutagenesis. *Ann Rev Pharmacol Toxicol*. 1990, 30, 91-121.
- 6 Robinson, JH et al. A recombinant chimeric enzyme with a novel mechanism of action leading to greater potency and selectivity than tissue-type plasminogen activator. *Circulation*, 1992, 86, 548-552.

Continued on page 70

Therapeutics

Figure 1

Domain shuffling with added chemistry: outline structure of a hypothetical hybrid protein therapeutic incorporating two effector domains and a pro-inflammatory pharmacophore mutated to improve therapeutic ratio and post-translationally modified to improve biodistribution



Continued from page 69

7 Moreland, LW et al. Phase I/II trial of recombinant methionyl human tumour necrosis factor binding protein PEGylated dimer in patients with active refractory rheumatoid arthritis. *J Rheumatol.*, 2000, 27, 601-609.

8 Linton, SM et al. Therapeutic efficacy of a novel membrane-targeted complement regulator in antigen-induced arthritis in the rat. *Arthritis & Rheumatism*, 2000, 43, 2590-2597.

9 Dong, J et al. Strategies for targeting complement inhibitors in ischaemia/reperfusion injury. *Molecular Immunology*, 1999, 36, 957-963.

10 Koechendoerfer, GC et al. Total chemical synthesis of the integral membrane protein Influenza A Virus M2: role of its C-terminal domain in tetramer assembly. *Biochemistry*, 1999, 38, 11905-11913.

Conclusion

A recent survey of protein therapeutics suggested that of about 250 non-vaccine agents in development in the US alone, only around 15 exploited one or more of the combinatorial, domain engineering or post-translational modification strategies noted above. This suggests that the full potential for these approaches is far from fully realised, particularly in view of the fact that domain function assignments in the human genome are in their infancy. Since they offer ways of driving biological insight into protein design and circumventing the crude 'clone and own' approach to patenting, there may also be good intellectual property reasons for adopting the domain-based, screening-bounded and chemistry-refined approaches to development of much greater

numbers of 'smart' protein-based pharmaceuticals. In this article, I have not mentioned monoclonal antibodies and hardly touched on vaccines. Those two omissions alone will, I hope, give some context to the vast untapped combinatorial therapeutic potential of the polypeptide chain. **DDW**

Dr Richard Smith received his doctorate from Oxford University for work on photaffinity labelling. Within the Beecham group and later SmithKline Beecham, he worked on applied enzymology and protein chemistry especially for the development of thrombolytic agents. He was a co-founder of the biopharmaceutical company Adprotech Ltd and is currently its Chief Scientific Officer.

ADVERTISEMENT INDEX

Accelrys	51	Gilson Inc	IFC,3	MIPTec	58
Arque Inc	53	Hamilton Bonaduz AG	29,31	PanVera Corporation	36
Biofocus plc	18	IBC Europe	55	Perkin Elmer Life Sciences	IBC
Biostreet.com	14	IBC USA	42	Phase-I Molecular Toxicology Inc	21
Cambridge BioScience Ltd	17	Incyte Genomics Inc	OBC	Physiome Sciences Inc	46
Cellomics Inc	39	Informax Inc	4	Promega Corporation	45
Chemnavigator.com	10	Kelly Scientific Resources	6	Thermo Finnigan Corporation	64
Chromagen Inc	40	Luminex Corporation	33	Tripes Inc	60
Corning Life Sciences	43	MDS Proteomics	67	Upstate Discovery Ltd	25
ECPI	68	Medichem Research Inc	26	Xanthon	34
Gene Logic Inc	8,9	Micromass Ltd	57		