a natural product of the eye and because it has potent anti-angiogenic activity. PEDF is a major angiogenic inhibitor in the vitreous and is probably responsible for the normal avascularity of several of the ocular tissues. Loss of PEDF has a 'permissive role' in ischemia-induced neovascular growth [5]. Therefore, gene therapy by delivery of PEDF to suppress ocular neovascularization could be more promising.

Although several different gene products can cause inhibition of ocular neovascularization, only partial suppression was observed in most of these cases. For example, blocking VEGF inhibits ~50% of retinal neovascularization. Restoration of the balance by provision of endogenous inhibitors, such as PEDF, and blocking of stimulators, such as VEGF, at the same time to arrest the progression of ocular neovascularization, might have synergistic effects.

Anti-angiogenic gene therapy has not been tested in clinical trials. All of the gene therapies in clinical trials against cancer target the tumour cell; the major obstacles include heterogeneity and drug resistance of tumour cells. Anti-angiogenic gene therapy directly targets the genetically stable endothelial cells, does not encounter drug-resistant tumour cells, and inhibits tumour growth independently of tumour cell heterogeneity in addition to its prolonged effect [6]. Therefore, it might have potential as a feasible and effective strategy in the treatment of cancer, ocular disorders and other angiogenesis-related diseases.

References

1 Bussolino, F. *et al.* (1997) Molecular mechanisms of blood vessel formation. *Trends Biochem. Sci.* 22, 251–256
2 Jimenez, B. and Voipert, O.V. (2001) Mechanistic insights on the inhibition of tumor angiogenesis. *J. Mol. Med.* 78, 663–672
3 Gao, G. *et al.* (2001) Unbalanced expression of VEGF and PEDF in ischemia-induced retinal neovascularization. *FEBS Lett.* 489, 270–276
4 Rasmussen, H.S. *et al.* (2001) Looking into anti-angiogenic gene therapies for disorders of the eye. *Drug Discov. Today* 6, 1171–1175
5 Chader, G.J. (2001) PEDF: Raising both hopes and questions in controlling angiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 98, 2122–2124
6 Cao, Y. (2001) Endogenous angiogenesis inhibitors and their therapeutic implications. *Int. J. Biochem. Cell. B.* 33, 357–369

Guoquan Gao and Jianxing Ma
Department of Ophthalmology,
Medical University of South Carolina
167 Ashley Ave, Charleston, SC 29403, USA
tel: +1 843 792 9180
fax: +1 843 792 1723

Taking the complexity out of protein sequence analysis ▼

The importance of revealing the function and properties of novel protein sequences has recently become evident to the international scientific community [1]. Both *in vitro* and *in silico* techniques are required for discovering these new protein sequences, their roles in the cell and their potential as drug targets. Eventually, a synergistic effect between *in silico* and *in vitro* approaches will facilitate the synthesis of novel and more specific drugs, leading to increased longevity and better health for the human population.

The availability of interactive sequence databases on the Internet has become important during the past decade. Sequence databases, such as Entrez GeneBank [2] and SWISSPROT [3], provide genomic sequences for many organisms, such as *Homo sapiens*, *Drosophilia melanogaster*, *Caenorhabditis elegans* and *Mus musculus*. To discover the properties of this vast number of protein sequences, from a phylogenetic perspective, as well as a proteomic one, several interactive tools have gradually emerged. Examples of these can be found at http://www.expasy.org, which provides and cross-links a large collection of tools for predicting translation, secondary structure, transmembrane regions, glycolysis patterns and other protein properties. Discovering what is most and least relevant for determining the crucial properties of an unknown protein sequence is the goal. However, it can be quite difficult because of the large number of interactive tools. Courses, personal preferences or advice from colleagues are usually the way we 'get used to' using the same engines time and again for analyzing protein sequences. However, this population of engines grows every year, and keeping updated with the various possibilities is important.

Therefore, an easy step-by-step procedure has been supplied (Figs 1 and 2), to delineate the possibilities and availabilities of different types of proteomic engines that are available on the Internet. Many of the listed tools are based on BLAST algorithms [4], such as ProDom [5], the Protein Information engine [6] and the Protein Fingerprint Database [7]. Others additionally apply Hidden Markov Models statistical analysis to sequence–structure correlations [8], as used in the SAM-99 engine [9] and HMM Superfamily engine [10]. Additionally, an interactive interface of the CLUSTALW package [11] has been suggested because of its frequent use in making multiple sequence alignments. Furthermore, the Prosal-interface, which connects several secondary structure prediction-, homology modeling-, threading- and sequence alignment-interfaces into one submission form, is proposed.

The order for the use of these engines is presented to facilitate the decomposition of pre/pro-regions (functionally superfluous regions removed after secretion from Golgi apparatus) down to catalytic domains, hormone-binding domains and other core parts of important physiological function. Once these sub-components of the full sequences have been identified, and superfluous regions have been removed, the suggested modeling procedures become much easier to perform and the results are more reliable

1359-6446/02/$ – see front matter PII: S1359-6446(02)02161-4

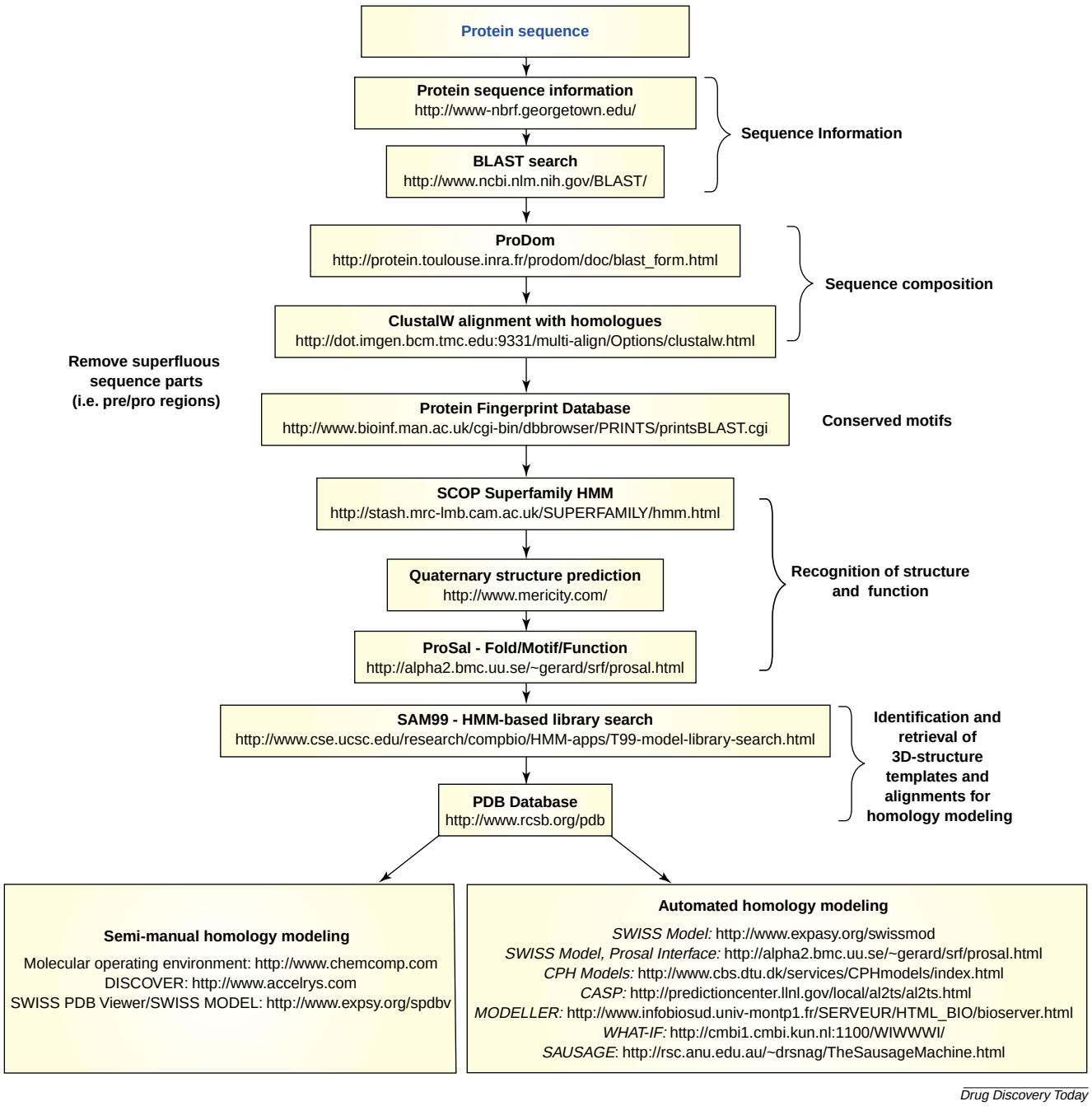

Figure 1. Condensing a raw sequence to functional data; fold, function, motifs and structure are recognized quickly with the right combination of interactive proteomic engines.

because of the increase in homologous templates, for example, to the catalytic domain.

Automated homology modeling

The simplicity of the proposed automated homology-modeling tools is quite similar to the sequence analysis tools. The SWISS MODEL program [12], available through http://www.expasy.org and the Prosal interface, the threading-based tertiary-structure prediction engine, SAUSAGE [13], and the modeling interface CPH models [14], all function by a simple submission of the sequence, to produce a PDB structure file sent via e-mail. The WHAT IF interface [15] is also suggested but requires the sequences to be submitted in a different format (PIR format) with an additional alignment, which can be generated through a sequence format translator (http://bioweb.pasteur.fr/seqanal/interfaces/fmtseq.html). MODELLER [16]

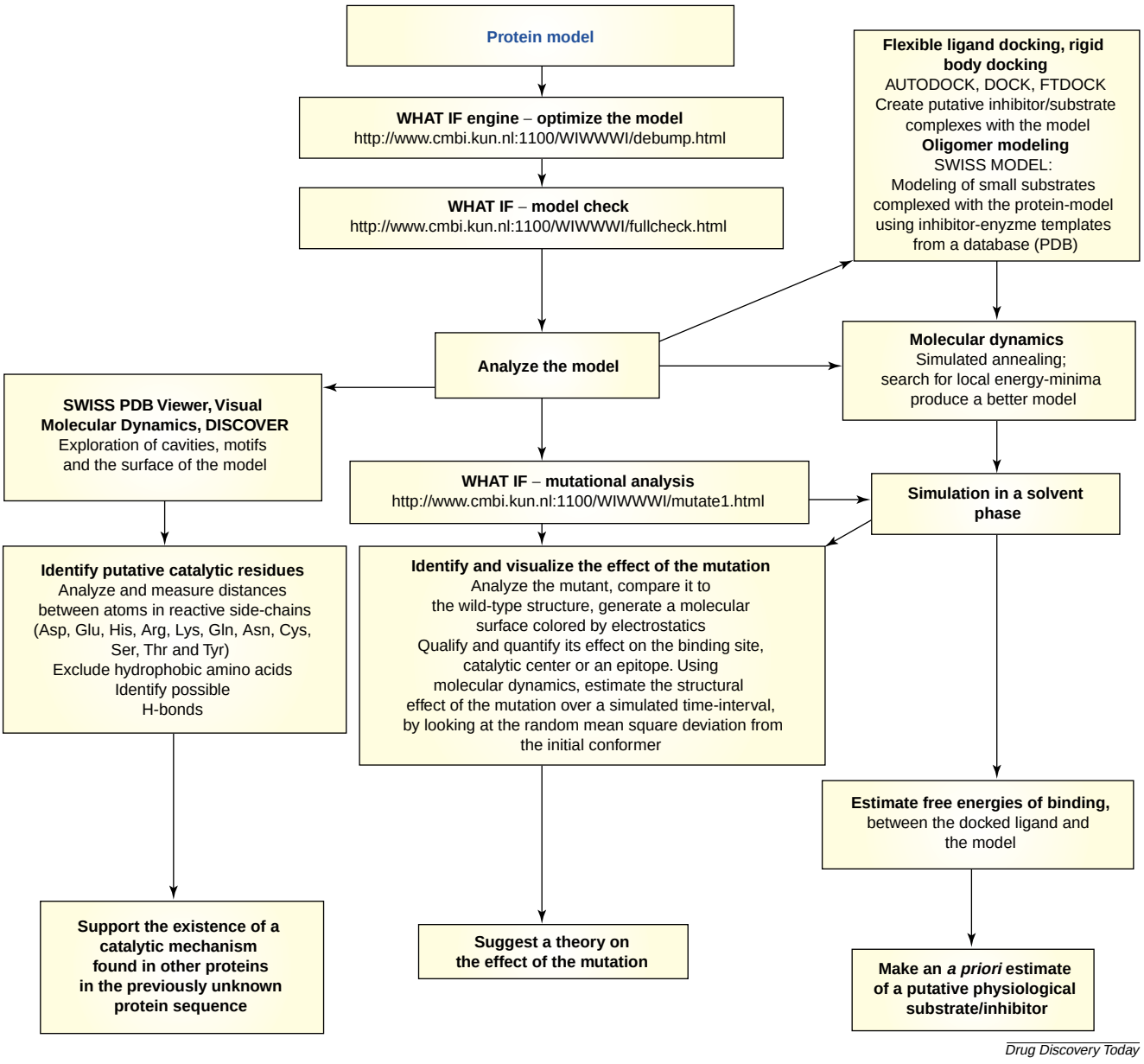

Figure 2. *In silico* possibilities to determine chemical and dynamical data of a protein structure; cavities, catalytic sites, substrate-binding regions, effects of mutations, and so on.

can be run interactively through the web-interface of the Centre of Structural Biochemistry (Montpellier, France; http://www.infobiosud.univ-montp1.fr/SERVEUR/HTML_BIO/bioserver.html). However, it requires the generation of input files using accessory interactive programs, which might be more time-consuming for new users than the alternatives suggested.

Once the model has been retrieved, a structure refinement can rapidly be performed through the WHAT IF interface, which optimizes geometric clashes between amino-acid side chains (energy minimization). The protein model can be viewed using the SWISS PDB Viewer program [12], which offers a wide variety of functions, such as molecular surface generation with electrostatic coloring, structure refinement, and so on. Visual Molecular Dynamics [17] is another option, which also offers the possibility of generating animations of a protein structure.

On the analysis level, many issues are required to classify the model; however, one that could provide important information is to find the closest structural neighbors of the finalized model. Holm and Sander [18] published the DALI-interface, where the PDB file

can be submitted in text format for a multiple three-dimensional structure superimposition session against a set of crystal structures in the FSSP-database (Fold classification based on Structure–Structure alignment of Proteins, at the European Bioinformatics Institute). This yields a brief report on the closest structural neighbors, which is crucial to confirm the function of the protein. (This procedure is better suited for models that are simulated and/or refined in a molecular dynamics session, because homology modeled structures tend to be most similar to their primary template.)

Other analyses many might wish to perform are to map the possible effects of mutations on the structure and function of the protein; these can be performed using the WHAT IF engine in a user-friendly way. Alternatively, SWISS PDB Viewer can be run from a PC/MAC, with more choices on the analysis of the mutation.

Novel catalytic residues

Discovering novel catalytic residues in a model that derives from a structurally, but not enzymatically, characterized group of proteins solely on an *in silico* basis is quite difficult. However, the observation of polar and/or charged residues in catalytic clefts, surrounded by hydrophobic and amphiphilic residues, can often be an important discovery, if the model is built on a set of 'good' templates. The hydrophobic residues around an active site usually have the role of lowering the dielectric constant in the catalytic center, which facilitates the transfer of electrons from donors to acceptors [19]. Therefore, the preliminary identification of putative residues that are involved in the catalytic mechanisms can be reduced to focus on hydrophilic and amphiphilic residues located in a cleft, because 75% of catalytic sites are located in cavities.

Alternatives to this procedure are, of course, applicable and more efforts to publish such user-friendly methodologies are required to supply the population of scientists working strictly from an *in vitro* aspect, with *in silico* methodologies to improve and increase the pace of achieving results.

Acknowledgements

I would like to thank Jonathan Harris from the Centre of Molecular Biotechnology, Queensland University of Technology (Brisbane, Queensland, Australia) for suggesting to publish this article.

References

1 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
2 Benson, D.A. *et al.* (2000) Genebank. *Nucleic Acids Res.* 28, 15–18
3 Bairoch, A. and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45–48
4 Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
5 Corpet, F. *et al.* (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res.* 28, 267–269
6 Barker, W.C. *et al.* (2001) Protein Information Resource: a community resource for expert annotation. *Nucleic Acids Res.* 29, 29–32
7 Wright, W. *et al.* (1999) BLAST PRINTS-alternative perspectives on sequence similarity. *Bioinformatics* 15, 523–524
8 Bystroff, C. *et al.* (2000) HMMSTR: a Hidden Markov Model for Local Sequence–Structure Correlations in Proteins. *J. Mol. Biol.* 301, 173–190
9 Krogh, A. *et al.* (1994) Hidden Markov models in computational biology: Applications to protein modeling. *J. Mol. Biol.* 235, 1501–1531
10 Gough, J. *et al.* (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* 313, 903–919
11 Thompson, J.D. *et al.* (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680
12 Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723
13 Huber, T. *et al.* (1999) Sausage: protein threading with flexible force fields. *Bioinformatics* 15, 1064–1065
14 Lund, O. *et al.* (1997) Protein distance constraints predicted by neural networks and probability density functions. *Protein Eng.* 10, 1241–1248
15 Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8, 52–56
16 Sali, A. *et al.* (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins* 23, 318–326
17 Humphrey, W. *et al.* (1996) VMD – Visual Molecular Dynamics. *J. Mol. Graph.* 14, 33–38
18 Holm, L. and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* 25, 231–234
19 Fersht, A. (1985) *Enzyme Structure and Mechanism*. (2nd edn), Freeman, W.H. & Co.

Sergio Manzetti
Proinformatix.com
Lippestad
1825 Tomter
Norway