

Structural biology and diabetes mellitus: molecular pathogenesis and rational drug design

Work group chairman: **T. Blundell**¹

Work group co-chairman: **R. Hubbard**²

Report prepared by: **M. A. Weiss**³

¹ Molecular Biology Laboratory, Department of Crystallography, Birkbeck College, University of London, London, UK; ² Department of Chemistry, University of York, York, UK; ³ Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA

Summary. Emerging concepts in the aetiology and pathogenesis of Type 1 (insulin-dependent) diabetes mellitus may offer new opportunities for treatment and cure. Here we describe recent advances in structural molecular biology and molecular design relevant to rational drug discovery. Such approaches focus on the three-dimensional structures of macromolecules and their interactions. In the coming decade such techniques may be applied to a wide variety of diabetes-related targets.

Key words: X-ray crystallography, nuclear magnetic resonance, molecular modelling, rational drug design.

Introduction

Recent progress in delineating the molecular pathogenesis of diabetes mellitus [1–4] offers the exciting prospect of designing novel therapies based on analysis of macromolecular structures and their interactions. The cloning and sequencing of genes related to the insulin signalling pathway [5, 6] and to disease susceptibility (see "Treatment of Diabetes Mellitus") promise to provide many new targets for drug discovery. In the coming decade, the stage is thus set for the application of modern techniques of structural chemistry and computer modelling. The purpose of this working paper is to review these techniques, their strengths and limitations, and to suggest possible future applications.

This report is presented in two parts. In the first part, general methods of structural investigation are reviewed with emphasis on the roles of X-ray crystallography, nuclear magnetic resonance and molecular modelling. Where possible, examples of relevance to diabetes mellitus are cited. In the second part, complementary investigations of systems of broad interest to future developments in techniques of molecular design are presented. These span a range of disciplines, including enzymology, immunology, virology, and vaccine development.

Protein structure and design

Molecular dissection of biochemical mechanisms and rational design of new drugs require knowledge of the three-dimensional structures of macromolecules and their interactions. In recent years, recombinant DNA technology has provided not only the sequences of many proteins that exist in minute quantities in the cell, but has also provided expression systems in which large quantities can be produced in the laboratory for biophysical studies. Study of the increasing number of these structures defined by X-ray analysis and nuclear magnetic resonance (NMR) provides insights into design of novel molecules that might be useful drugs. Although screening for active molecules continues to be widely used, there is an increasing interest in rational approaches. These usually involve the description of the three-dimensional structure of a protein or DNA complexed with a ligand, followed by exploration of chemical modifications of either the macromolecule or the ligand. This approach poses two major challenges: first, to define the three-dimensional structure experimentally or by prediction, and second to derive methods and rules for designing active molecules based on this structural information (Fig. 1). There are presently four major approaches used to obtain such knowledge: X-ray diffraction, NMR spectroscopy, knowledge-based protein folding, and computer-based ligand design.

X-ray diffraction. Crystallisation has long provided a convenient route to the purification of insulin for use in the treatment of diabetes, and different crystalline forms have had important application in the production of clinical preparations. Thus it was insulin crystals that attracted Hodgkin's attention at an early stage in the history of protein chemistry, leading to pioneering studies by X-ray diffraction in the 1930s. The successful structural analysis, completed 35 years later [7, 8], stimulated an increased awareness in the diabetes research community of the wealth of information contained in the complex three-

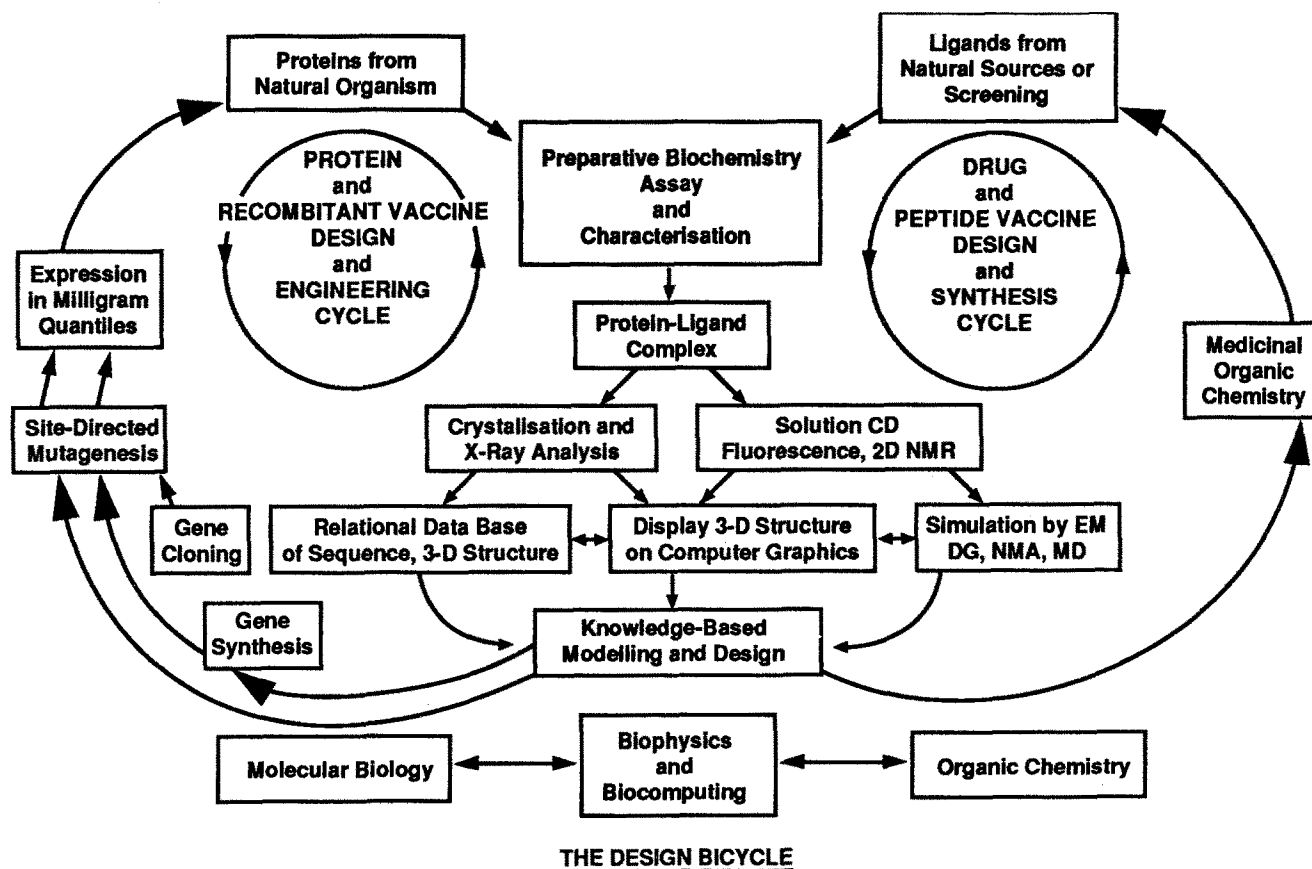


Fig. 1. Flow chart of interdisciplinary steps in the design of novel proteins and drugs. The variety of tasks involved may be organised as two interlocking cycles (the "Design Bicycle"): protein and recombinant vaccine design and engineering (left), and drug and peptide vaccine design and synthesis (right)

dimensional structures of proteins. In the ensuing years, attention has focussed on the need to understand hormone-receptor interactions [9, 10], mechanisms of signal transduction [5], and "downstream" events in the post-receptor signalling pathway [6]. Although these molecules are under intensive current study, to date no new crystal structures of such proteins have been determined.

Over the past years there have been many advances in protein crystallography that derive from advances in X-ray source and detection technology. Synchrotron sources, for example, provide intense, monochromatic and tuneable beams of X-rays; these can provide splendid diffraction patterns from the smallest crystals of the largest macromolecules. There have been advances in detectors so that the diffraction images, which were originally measured on photographic film or obtained reflection by reflection with counters, can now be measured in a few hours (or even minutes) for most proteins. Large computers have also expedited the analysis. However, there are still major hurdles in the way of many macromolecular analyses. These are primarily in the area of crystallisation, particularly of complex multidomain proteins such as receptors and transducers of hormone and growth factor signals. Such molecules are present in very small quantities, are often post-translationally modified by glycation or phosphorylation, and usually are comprised of flexibly linked domains. However, recombinant DNA

techniques make it possible to express single domains, which may be more homogeneous and hence more likely to crystallise [11]. This seems a promising approach for the transmembrane receptors of insulin, insulin-like growth factor and many other growth factors, where the best chance for obtaining crystals suitable for X-ray analysis may come from the study of the isolated extracellular (ligand-binding) domains or cytosolic tyrosyl kinases.

NMR spectroscopy. NMR methods are based on observation of nuclear spins and their interactions in a molecule. Although the first NMR resonance was observed in $^7\text{Li}^+$ beams in the 1930s, it was not until the 1980s that this technique was shown to provide a usable approach for determining the three-dimensional structures of proteins. Key breakthroughs were made by Ernst et al. [12] in the theory of multidimensional NMR spectroscopy and by Wuthrich and co-workers [13] in its systematic application in biochemistry. NMR differs from X-ray crystallography in its ability to obtain structural information in solution. Its range of application has recently been extended by three-dimensional and four-dimensional hetero-nuclear methods [14] to proteins whose molecular weights are up to 30-40 kDa.

In the past 5 years, NMR methods have been used to characterise a number of proteins of biomedical interest, including epidermal growth factor, nuclear receptors for

steroid hormones, and homeodomains involved in tissue-specific gene expression and development. In these studies, three-dimensional models are calculated in accord with experimental restraints using distance-geometry and restrained molecular dynamics methods. Direct comparisons of crystal and NMR structures have demonstrated the consistency of these two methods for stably folded domains. In addition, NMR studies have provided complementary dynamic information, which may be of central importance in flexible systems. In addition to direct methods for NMR structure determination, indirect methods have been developed to characterise the bound-state structure of ligands [15].

The structure of insulin has recently been determined in solution [16] (Fig. 2), and has interesting implications for the mechanism of receptor binding. The NMR data address two fundamental questions that were raised by crystal structures. The first is posed by the observation of multiple different insulin conformations among different crystal forms [9]. Which form best represents the insulin structure in solution? The NMR data strongly suggests that the insulin monomer is flexible [17-19]. Its detailed structure at any one time may differ from any one crystal form, but on average appears to span the range of different (T-state) crystal

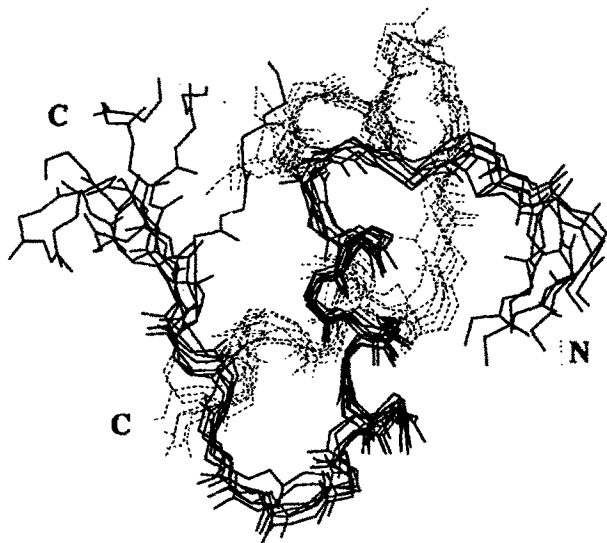


Fig. 2. Solution structure of an engineered insulin monomer, as determined by 2D-NMR methods and distance-geometry/simulated annealing. The B-chain is shown in solid line, and the A-chain in dashed line. The analogue contains three mutations (HisB10 Asp; ProB28 Lys; and LysB29 Pro) that destabilise the hexamer- and dimer-forming surfaces without loss of bioactivity

forms [20]. Which form best represents the active state of the hormone? Analysis of "anomalous" mutant human insulins — i.e. analogues whose bioactivities are not readily rationalised from crystal models — suggests that insulin undergoes a change in conformation upon binding to the insulin receptor. Complementary crystallographic and NMR studies of active and inactive analogues support the hypothesis that the C-terminal region of the B-chain

(residues B24-B30) moves to expose the underlying hydrophobic surface of the A-chain (residues A1-A3) for receptor contact [16].

Protein modelling. In pioneering studies of the refolding of small globular proteins, Anfinsen and co-workers introduced in the early 1960s the concept of a protein-folding code relating linear sequence to structure. The chemical basis of protein folding remains today a major unsolved problem in molecular biology. Although recent research on protein folding has shown that chaperons play an important role in protein folding, particularly of oligomeric proteins, it is generally accepted that the sequences of most proteins contain the information required to define their three-dimensional structures. In principle, this implies that computer simulations of the folding process should allow the definition of the three-dimensional structure of any protein whose sequence has been defined. In fact, the inadequacy of atomic potentials, combined with the complexity of the simulations (which challenges even the most powerful computers), means that "de novo" prediction of structure will not be a viable proposition in the immediate future.

An alternative approach has been developed that exploits the fact that three-dimensional structures of proteins appear to belong to a limited number of "families," each characterised by a common fold [21-23]. The existence of such families motivates a search for a solution to the "inverse" folding problem: if we know the three-dimensional structure of a protein, can we predict all those sequences that adopt this protein fold [24]? If a newly defined sequence can be associated with a protein fold that has been characterised experimentally, this can be used to gain an insight into the fold that the new sequence adopts. The inverse folding problem can be approached in three stages. The first of these generates all the sequences that can adopt a known three-dimensional structure. The second compares the sequence of the "unknown" with the sequences or templates that characterise each known three-dimensional structure. The third stage takes the sequence together with the known three-dimensional structure and generates a model of the protein [23].

In recent years, there has been remarkable progress in methods that generate templates for sequences that can adopt a known three-dimensional structure. Environment-dependent substitution tables [25-28] can define the amino acids that can occupy any position in a known three-dimensional structure. This has allowed the derivation of templates that can recognise distantly related sequences of any known three-dimensional fold. In principle, this implies that sequences of proteins that adopt a known three-dimensional fold can be recognised. This may provide a route to modelling many of the receptors and enzymes that are involved in the complex pathways that transduce the signals of insulin and other hormones and growth factors involved in diabetes.

A modelling procedure has been developed based on homologous (divergently evolved) or analogous (convergently evolved) structures. Most such methods depend on constructing a model from fragments of known structures. The "framework," or common structure of the fold, is taken from other members of the protein family identified in the template search. The variable regions can be selected from the wider data base of known three-dimensional structures. Finally, sidechains can be modelled from rules obtained by comparing families of known three-dimensional structures. Such methods [23] provide good models of closely homologous structures and in favourable cases can be used to build useful three-dimensional models of many proteins. For example, protein kinases such as the insulin receptor tyrosyl kinase that are involved in the intracellular transduction of hormone signals can be constructed on the basis of the experimentally defined structure of cAMP-dependent protein kinase [29].

Ligand design. Understanding the interaction between a ligand and its target protein is a general problem in physical and bioorganic chemistry which is becoming increasingly important in rational drug design as the three-dimensional structures of further targets for therapeutic design, such as key metabolic and processing enzymes and of signalling proteins, are determined. Nevertheless, theoretical methods of rational drug designs are far from perfect. In considering underlying chemical principles, it is important to distinguish between the design requirements of agonists and antagonists. The former must be precise, reflecting specific geometric features of the active site. In contrast, design requirements for an antagonist or inhibitor are often less exact, since non-native modes of binding may be effective in preventing substrate binding or allosteric activation. Thus, whereas modelling of agonists may require a high-resolution representation of the molecular force field, more approximate calculations may be sufficient for inhibitor design.

Of special interest are design problems posed by flexible ligands, such as peptides. Such ligands have many degrees of freedom, and their potential energy surfaces are likely to have multiple minima. Simple empirical or semiclassical energy-minimisation procedures are unlikely to locate the preferred conformation or mode of binding. Modelling studies require the inclusion of experimental data, such as may be provided by structural studies of analogous ligand-protein complexes. In the absence of such structures, valuable insight into the bound-state structure of the ligand may be obtained from transferred nuclear overhauser effect experiments [15]. The problem of flexible ligands may also be circumvented through study of conformationally restrained analogues.

Systematic analysis of a series of related ligands may enable statistical correlations to be drawn between physicochemical features and bioactivity; such correlations are typically cross-validated using analogues not

represented in the training set and in favourable cases may be used to evaluate lead structures for further analogue design. This approach may be used in the absence of structural knowledge of the protein-active site and extended to enable a quantitative analysis of structure-activity relationships by constraining the positions of putative functional groups in the protein [30].

Applications and future developments

The above approaches to the determination of molecular structures share a common reductionist theme: the simplification of supramolecular biological systems into component parts for individual analysis. An outstanding example is the dissection of the antibody into a series of successive subdomains: F_{ab} , F_v , and single-chain F_v . This strategy has permitted detailed focus on the structural principles of hapten specificity [31-33] with application to the design of novel antibody combining sites [34]. In a broad sense the reductionist approach is supported by the modular construction of proteins and their proposed evolution by exon shuffling. It is important to recognise, however, that this theme has the intrinsic limitation of decoupling the component parts of an integrated system. Lost in such a reduction will be any overarching organisational or regulatory principles that govern the operation of the system as a whole. For this reason, it is of critical importance to test the predictions of molecular design in a meaningful biological system. In the following subsections, we consider specific examples of molecular design in systems of medical interest.

Immune recognition and the antibody combining site. Structural characterisation of the antibody combining site (ACS) has provided important insight into the interaction between a paratope and epitope, into structural variations among immunoglobulin classes, and into the structural implications of somatic variation in the maturation of the antibody response. Ten crystal structures have to date been determined for protein- F_{ab} , protein- F_v , peptide- F_{ab} , and hapten- F_{ab} complexes [31-33]. These structures, which exhibit both general similarities and particular differences, provide a foundation for the design of novel combining sites that can act as therapeutic and catalytic reagents.

Of recent interest are two complementary sets of investigations. The first focuses on comparing a series of antibody-protein complexes. Using the avian lysozymes as a model system, Poljak, Davies and their respective co-workers [31, 32] have investigated the nature of an epitope on an intact protein and the structural basis of affinity and specificity. An idio-anti-idio complex has also been determined and is of general interest in relation to the "internal image" hypothesis. It was found that the anti-idio ACS does not provide such an image of the parent epitope [35]. In particular, changes in sidechain configurations in the interface are observed that

accommodate detailed topographical differences between the two molecular surfaces. These results make it unlikely that anti-idiotypic antibodies may find ready application as an immunogen in vaccine design.

The structures of the protein-antibody complexes have general implications for mechanisms of protein-protein recognition. Although the interface between the protein and ACS resembles in several respects that between an enzyme and peptide inhibitor (e.g. trypsin-trypsin inhibitor), the ACS differs in the larger proportion of aromatic contact residues (40-50%) and smaller number of mainchain contacts. This difference may reflect the requirement in the protease inhibitors for the proper presentation of the scissile bond [36]; in contrast, the ACS has presumably evolved to be highly adaptable to confront an ever-changing immunological challenge. We speculate that aromatic residues may be selected in the ACS since they provide (i) a hydrophobic surface for binding, (ii) an asymmetric distribution of electronic and nuclear charges for weak polar interactions, and (iii) in the case of tyrosine, histidine, and tryptophan, a specific pattern of hydrogen-bond donors and acceptors. These diverse chemical interactions are indeed seen in the co-crystal structures [31-33, 35].

The engineering of antibody active sites to endow them with catalytic activity provides an intriguing approach to producing novel enzymes. Such "abzymes" may contain metal binding sites. Metal-binding sites are often seen in the active sites of enzymes, so this design goal represents a promising first step toward the rational construction of catalytic antibodies. Tainer and colleagues [34] have described the remodelling of complementary-determining regions to contain tetrahedral coordination sites for copper and zinc. Although crystal structures of these engineered F_v fragments have not been obtained, metal-binding is conferred by the designed changes and exhibits the expected dependency on the identity and spacing of amino-acid substitutions in the putative histidine ligands.

Novel enzymes may also be engineered by modifying existing enzymes. For example, human [Cu, Zn]-superoxide dismutase (SOD) protects cells against oxidative damage in a reaction of pharmacologic interest. SOD is a "perfect" enzyme in the sense that the rate of catalysis is limited by diffusion. Of particular interest is a possible role for electrostatic guidance in biasing the random-walk of the substrate into the active site. In the crystal structure of SOD, a conserved network of hydrogen-bonded residues above a deep active-site cleft engenders an electrostatic field that is predicted to attract the negatively charged superoxide molecule. To test this hypothesis, site-directed SOD mutants were designed based on electrostatic calculations. The mutant proteins were experimentally demonstrated to exhibit an approximately three-fold enhancement in the rate of catalysis. The observed dependence of the reaction rate on ionic strength (i.e. due to changes in electrostatic shielding) was rationalised by Brownian dynamics simulations [37].

Applications of ligand design in molecular pharmacology. We describe three biological systems that provide interesting examples of the promise and difficulties of rational ligand design. The first, inhibitors of angiotensin-converting enzyme, represents a landmark in rational drug design [38] with widespread clinical importance in the control of hypertension [39]. Angiotensin converting enzyme (ACE) is a member of an extended family of Zn-metalloenzymes. Captopril (Squibb) and enalapril (Merck) were designed as selective ACE inhibitors and have widespread clinical use. In addition, there is a large set of congeners, which provides a unique data base for the development and testing of ligand-design algorithms. An important simplification in this system is the structural requirement of zinc coordination in the protein active site. Such coordination provides a common frame of reference for the alignment of functional groups in a candidate compound in pseudo-receptor calculations [30].

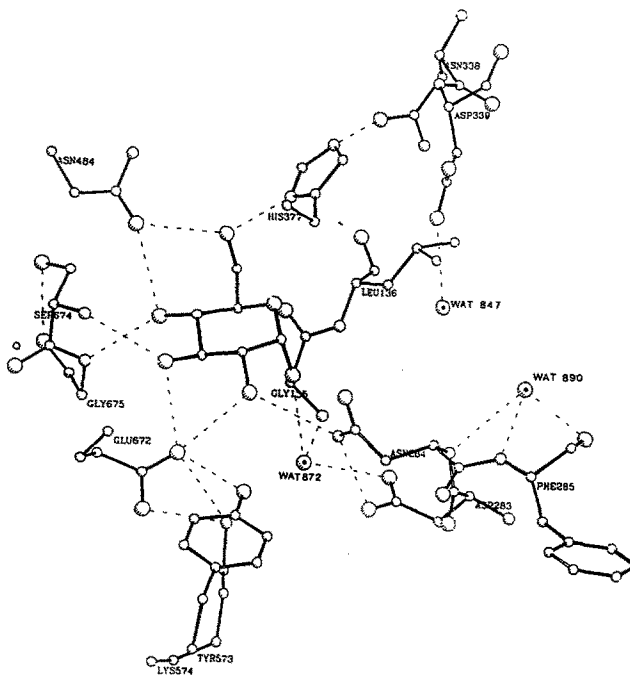


Fig. 3. Contacts formed between glucose and glycogen phosphorylase. The structure of this complex was used as the basis for designing glucose-derived inhibitors of glycogen phosphorylase (taken from Martin et al. [41])

The second example is that of glycogen phosphorylase, which has been investigated in detail by X-ray crystallography and computational chemistry [40, 41]. Potent inhibitors of this enzyme, which plays a key role in blood glucose regulation, might be useful in the treatment of diabetes. The three-dimensional X-ray structure of T-state rabbit muscle glycogen phosphorylase b (which has 80% sequence identity to the human liver enzyme) has been solved to high resolution. The complex between this form of the enzyme and glucose, a physiological inhibitor, has also been determined [40]. The X-ray structure of the glucose-enzyme complex was used to model glucose derivatives substituted at the C1 position. Figure 3 shows

the structure of the glucose-glycogen phosphorylase complex at the region of the catalytic site of the enzyme [41]. Modified substituents at C1 were modelled to interact with additional favourable binding sites within the catalytic site, identified using program GRID [42]. A number of these designed compounds, predicted to adopt low energy binding conformations, were synthesised and subsequently tested for their ability to inhibit glycogen phosphorylase. X-ray analysis was also performed to study how the compounds interact with the catalytic site of the enzyme. The observed binding modes of the compounds were generally close to those predicted, but the binding affinity was not better than that measured for glucose. However, the results did provide guidelines for the design of more potent inhibitors which will be modelled, synthesised and studied in the future. A final example is provided by enzyme-activated anti-metabolites as anti-viral and anti-neoplastic agents. This important area has recently been stimulated by the cloning and purification of target enzymes. We will focus on the herpes simplex virus (HSV) thymidine kinase (tk) and the use of pseudosubstrates such as acyclovir to interrupt the viral life cycle [43-46]. This system has the following features: (i) The enzyme is of unknown structure but its sequence belongs to a well-recognised family of enzymes (nucleoside-binding

enzymes). X-ray structures have been determined for distantly related family members, including adenylate kinase and elongation factor TU. Modelling the HSV-tk structure thus poses an important test of knowledge-based prediction methods that may be tested experimentally. (ii) A parent pseudosubstrate is known (acyclovir) and widely used to treat HSV infections in immuno-compromised hosts. Acyclovir thus provides a lead structure for analogue design. In particular, the open-chain sugar moiety of acyclovir provides a model of a flexible ligand whose bound-state structure may be investigated through the design of conformationally constrained analogues. (iii) Synthesis of the above knowledge-based protein prediction and ligand-based pseudoreceptor calculations may permit a detailed model to be constructed of the interaction between pseudosubstrates and the active site (Fig. 4). In the absence of a crystal structure for direct verification, this model may be tested by site-directed mutagenesis. If successful, the convergence of these three approaches in the case of HSV-tk will provide a general strategy for rational drug design.

Conclusions and recommendations

Type 1 diabetes presents a broad range of challenges to drug design, reflecting the complexity of its primary aetiology, pathogenesis, and long-term complications. Each of the disciplines reviewed in this volume — the immunological destruction of pancreatic Beta cells and possible prodromal intervention, the pharmacology of insulin preparation and delivery, sites of secondary injury in blood vessels, kidney, nerves, and eye — are associated with distinct molecular targets and mechanistic issues. How may an "immune antagonist" be designed to abort islet destruction and salvage insulin-secretory function? May further characterisation of the insulin receptor or post-receptor pathway permit the development of novel hypoglycaemic agents? Will it be possible to formulate strategies of therapy to avoid or reduce the incidence of long-term complications? The coming decade is likely to witness an interdisciplinary convergence of interest on these and related issues. This convergence will reflect continued advances in human genetics, immunology, molecular biology, structural chemistry, and techniques of computer-aided drug and protein design. This review is motivated by the new opportunities for therapeutic intervention that these advances promise. Of critical importance will be the application of the methods of X-ray crystallography, multi-dimensional NMR, and computer-based modelling. By defining structural mechanisms of Beta-cell development, regulation of insulin synthesis and release, and of metabolic regulation in target tissue, future progress in this "The Decade for the Cure" will be enhanced by creative synthesis of the clinical and basic sciences.

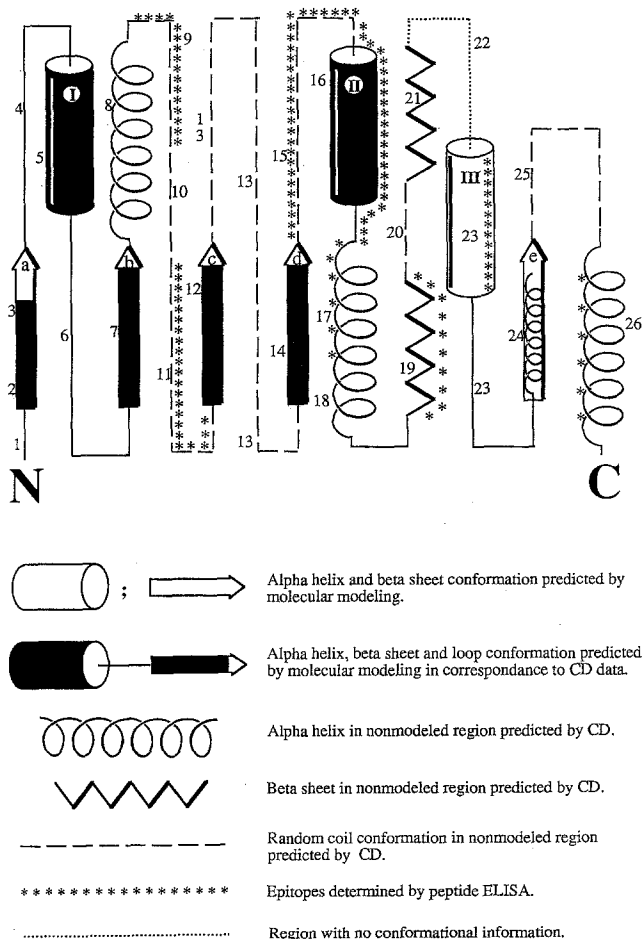


Fig. 4. Prediction of secondary structure in the herpes simplex virus thymidine kinase, as developed by Folkers and colleagues [46]

Acknowledgements. This report is the result of the Work Group 9: Molecular Design at the Third JDF World Conference on Diabetes in Monte Carlo, Monaco, 8-10 March 1992. Chairman: T. L. Blundell. Co-chairman: R. Hubbard. Work group participants: G. Folkers¹, E. D. Getzoff², C. Humblet³, G. R. Marshall⁴, J. Martin⁵, R. J. Poljak⁶, S. Wodak⁷. Rapporteur: M. A. Weiss.

¹Department of Pharmacy, ETH Zentrum, Zurich, Switzerland; ²Department of Molecular Biology, Scripps Research Institute, La Jolla, California, USA; ³Parke-Davis Pharmaceutical Research Division, Ann Arbor, Michigan, USA; ⁴Department of Pharmacology, Washington University School of Medicine, St. Louis, Missouri, USA; ⁵Laboratory of Molecular Biophysics, Howard Hughes Medical Institute, Rockefeller University, New York, New York, USA; ⁶Immunologie Structurale, Institut Pasteur, Paris, France; ⁷Unite de Conformation de Macromolecules Biologiques, Universite Libre de Bruxelles, Brussels, Belgium.

References

- Bell GI (1991) Lilly Lecture: molecular defects in diabetes mellitus. *Diabetes* 40: 413-422
- Bell GI, Xiang KS, Newman MV et al. (1991) Gene for non-insulin-dependent diabetes (maturity-onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. *Proc Natl Acad Sci USA* 88: 1484-1488
- Froguel P, Vaxillaire M, Sun F et al. (1992) Close linkage of glucokinase locus on chromosome 7p to early onset non-insulin-dependent diabetes mellitus. *Nature* 356: 162-164
- Faustman D, Li X, Lin HY et al. (1991) Linkage of faulty major histocompatibility complex class I to autoimmune diabetes. *Science* 254: 1756-1761
- Sun XJ, Rothenberg P, Kahn CR et al. (1991) Structure of the insulin receptor substrate (RS-) defines a unique signal transduction protein. *Nature* 352: 73-77
- Skolnik EY, Margolis B, Mohammadi M et al. (1991) Cloning of PI 3 kinase-associated p 85 utilising a novel method for expression/cloning of target proteins for receptor tyrosine kinases. *Cell* 65: 83-90
- Blundell TL, Cutfield JF, Cutfield SM et al. (1971) Atomic position in 2-zinc insulin crystals. *Nature* 231: 506-511
- Peking Insulin Structure Group (1971) Insulin's crystal structure at 2.5 Å resolution. *Peking Rev* 40: 11-16
- Baker EN, Blundell TL, Cutfield GS et al. (1988) The structure of 2-Zn pig insulin crystals at 1.5: a resolution. *Phil Trans R Soc B* 319: 389-456
- Mirmira R, Tager HS (1989) Role of the phenylalanine B24 sidechain in directing insulin interaction with its receptor: importance of mainchain conformation. *J Biol Chem* 264: 6349-6354
- De Vos AM, Ultsch M, Kossiakoff AA (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 255: 306-312
- Ernst RR, Bodenhausen G, Wokaun A (1987) Principles of nuclear magnetic resonance in one and two dimensions. Clarendon Press, Oxford
- Wagner G, Withrich K (1982) Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J Mol Biol* 155: 347-366
- Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance NMR spectroscopy. Application to calmodulin. *Biochem* 29: 4659-4667
- Clore GM, Gronenborn AM (1983) Theory of the time dependent transferred nuclear overhauser effect: applications to structural analysis of ligand-protein complexes in solution. *J Mag Reson* 53: 423-442
- Hua QX, Shoelson SE, Kochogyan M, Weiss MA (1991) Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* 354: 238-241
- Kline AD, Justice RM (1990) Complete sequence-specific 1H-NMR assignments for human insulin. *Biochemistry* 29: 2906-2913
- Boelens R, Ganadu ML, Verheyden P, Kaptein R (1990) Two-dimensional NMR studies on des-pentapeptide-insulin: proton resonance assignments and secondary analysis. *Eur J Biochem* 191: 147-153
- Weiss MA, Hua QX, Lynch CS, Frank BH, Shoelson SE (1991) Heteronuclear NMR studies of an engineered insulin monomer: characterisation of the receptor-binding surface by selective 2H and 13C isotopic labelling with application to protein design. *Biochemistry* 30: 7373-7389
- Hua QX, Kochogyan M, Weiss MA (1992) Structure and dynamics of des-pentapeptide insulin in solution: the molten globule hypothesis. *Proc Natl Acad Sci USA* 89: 2379-2383
- Rooman MJ, Wodak S (1988) Identification of predictive sequence motifs limited by protein structure data base size. *Nature* 335: 45-49
- Rooman MJ, Kocher JP, Wodak SJ (1991) Prediction of protein backbone conformation based on seven structure assignments. *J Mol Biol* 221: 961-979
- Sali A, Overington JP, Johnson MS, Blundell TL (1990) From comparisons of protein sequences and structures to protein modelling and design. *Trends Biochem Sci* 15: 235-240
- Novotny J, Bruccoleri R, Karplus M (1984) An analysis of incorrectly folded protein models: implications for structure predictions. *J Mol Biol* 177: 787-818
- Overington JP, Johnson MS, Sali A, Blundell TL (1990) Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc R Soc Lond Ser B* 241: 132-145
- Luthy R, McLachlan AD, Eisenberg DC (1991) Secondary structure-based profiles: use of structure-conserving scoring tables in searching protein sequence databases for structural similarities. *Proteins* 10: 229-239
- Bowie JV, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253: 164-170
- Overington J, Donnelly D, Johnson MS, Sali A, Blundell TL (1992) Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Protein Sci* 1: 216-226
- Knighton DR, Zheng J, Ten Eyck LF et al. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253: 407-413
- Mayer D, Naylor CB, Motoc I, Marshall GR (1987) A unique geometry of the active site of angiotensin-converting enzyme consistent with structure-activity studies. *J Comput Aided Mol Des* 1: 3-16
- Amit AG, Mariuzza RA, Phillips SE, Poljak RJ (1986) Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* 233: 747-753
- Padlan EA, Silverton EW, Sheriff S, Cohen GH, Smith-Gill SJ, Davies DR (1989) Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc Natl Acad Sci USA* 86: 5938-5942
- Bhat TN, Bentley GA, Fischmann TO, Boulot G, Poljak RJ (1990) Small rearrangements in structure of F_v and F_{ab} fragments of antibody D1.3 on antigen binding. *Nature* 347: 483-485
- Roberts VA, Iverson BL, Iverson SA et al. (1990) Antibody remodeling: a general solution to the design of a metal-coordination site in an antibody binding site. *Proc Natl Acad Sci USA* 87: 6654-6658
- Bentley GA, Boulot G, Riottot MM, Poljak RJ (1990) Three dimensional structure of an idiotope-anti-idiotope complex. *Nature* 348: 254-257

36. Veerapandian B, Cooper JB, Sali A et al. (1992) Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci* 1: 322-328
37. Getzoff ED, Cabelli DE, Fisher CL et al. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature* 358: 347-351
38. Cushman DW, Cheung HS, Sabo EF, Ondetti MA (1977) Design of potent competitive inhibitors of angiotensin-converting enzyme, carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochem* 16: 5484-5491
39. Ulm EH, Greenlee WJ (1989) Inhibitors of the renin-angiotensin system enzymes. In: Sandler M, Smith HJ (eds) *Design of enzyme inhibitors as drugs*. Oxford University Press, New York, pp 146-177
40. Martin JL, Johnson LN, Withers SG (1990) Comparison of the binding of glucose and glucose 1-phosphate derivatives to T-state glycogen phosphorylase b. *Biochem* 29: 10745-10757
41. Martin JL, Veluraja K, Ross K et al. (1991) Glucose analogue inhibitors of glycogen phosphorylase: the design of potential drugs for diabetes. *Biochem* 30: 10101-10116
42. Goodford PJ (1985) A computational procedure for determining energetically favourable binding sites on biologically important molecules. *J Med Chem* 28: 849-857
43. Folkers G, Krickl S, Trumpp S (1989) Localisation of the essential structure for binding of antiviral agents to thymidine kinase by studying sequence homologies. *Arch Pharm* 322: 409-413
44. Folkers G, Sakahara K, Schwabel W, Eger K (1989) The development of pharmacophore models for thymidine-kinase-dependent virostatic nucleoside analogues. *Arch Pharm* 322: 395-398
45. Folkers G, Trumpp-Kallmeyer S, Gutbrod O, Krickl S, Fetzer J, Keil GM (1991) Computer-aided active-site directed modeling of the herpes simplex virus 1 and human thymidine kinase. *J Comput Aided Mol Des* 5: 385-404
46. Zimmermann N, Beck-Sickinger AG, Folkers G, Krickl S, Muller I, Jung G (1991) Conformational and epitope mapping of herpes-simplex-virus-type 1 thymidine kinase using synthetic peptide segments. *Eur J Biochem* 200: 519-528