Instituto Juan March
de Estudios e Investigaciones

34

CENTRO DE REUNIONES
INTERNACIONALES SOBRE BIOLOGÍA

Workshop on

Computational Approaches in the Analysis and Engineering of Proteins

Organized by

F. X. Avilés, M. Billeter and E. Querol

T. Blundell
J. M. Carazo
C. Chothia
J. Devereux
K. A. Dill
D. Eisenberg
I. Fita
F. Gago
J. Greer
M. Karplus
M. Levitt
G. M. Maggiora
M. A. Navia
M. Orozco
A. Rey
C. Sander
M. J. Sippl
O. Tapia
J. Thornton
A. Valencia
W. F. van Gunsteren
S. J. Wodak

Instituto Juan March
de Estudios e Investigaciones

CENTRO DE REUNIONES INTERNACIONALES
SOBRE BIOLOGÍA

34

Workshop on

Computational Approaches in the Analysis and Engineering of Proteins

Organized by

F. X. Avilés, M. Billeter and E. Querol

T. Blundell
J. M. Carazo
C. Chothia
J. Devereux
K. A. Dill
D. Eisenberg
I. Fita
F. Gago
J. Greer
M. Karplus
M. Levitt
G. M. Maggiora
M. A. Navia
M. Orozco
A. Rey
C. Sander
M. J. Sippl
O. Tapia
J. Thornton
A. Valencia
W. F. van Gunsteren
S. J. Wodak

The lectures summarized in this publication were presented by their authors at a workshop held on the 14th through the 16th of November, 1994, at the Instituto Juan March.

Depósito legal: M. 342/1995
I. S. B. N.: 84-7919-534-7
Impresión: Ediciones Peninsular. Tomelloso, 27. 28026 Madrid.

INDEX

PROGRAMME

COMPUTATIONAL APPROACHES IN THE ANALYSIS AND ENGINEERING OF PROTEINS

Monday, November 14th, 1994

I. Prospects in Computational Molecular Biology
Chairpersons: T. Blundell & F.X. Avilés

C. Sander - Protein structure: 1D-2D-3D.

D. Eisenberg - 3D profiles for assessing, designing, and refining proteins.

W.F. van Gunsteren - Computer simulation of proteins, prospects and limitations.

Software demonstrations.

II. Data bases, prediction and modelling
Chairpersons: C. Chothia & S.J. Wodak

T. Blundell - Rule-based approaches to the prediction and modelling of proteins.

J. Thornton - Protein folds and their determinants.

J.M. Carazo - Information processing and the need of biological data bases: An overview and a case study on the 3D density data base.

J. Devereux - Advances in amino acid substitution matrices.

Software demonstrations.

TUESDAY, November 15th 1994

II. (Continuation).
Data bases, prediction and modelling
Chairpersons: D. Eisenberg & E. Querol

A. Valencia - Structural and functional information contained in protein families: Engineering of specificity.

A. Rey - A hierarchical simulation method of protein folding.

S.J. Wodak - Extracting information on the protein 3D structure and folding pathway from the amino-acid sequence.

M. J. Sippl - Knowledge based force fields in the analysis, recognition and computation of protein folds.

Software demonstrations.

III. Molecular recognition and drug design
Chairpersons: M. Karplus & G.M. Maggiora

J. Greer - Role of homology modeling in the design of novel pharmaceutical agents.

M. Orozco - Molecular dynamics and free energy perturbation studies of the interaction between drugs and proteins.

F. Gago - Toward a more rational design of enzyme inhibitors: Phospholipase A_2 as a test case.

M.A. Navia - Structure-based drug design of VX-478, a potent and orally-bioavailable inhibitor of the human immunodeficiency virus type 1 (HIV-1) protease.

Software demonstrations.

WEDNESDAY, November 16th, 1994

IV. Models for protein folding and function
Chairpersons: M. Levitt & M. Billeter

G.M. Maggiora - Application of fuzzy sets and fuzzy logic to problems in protein science.

C. Chothia - Sequence determinants of protein folds.

I. Fita - Limited folding in proteins.

O. Tapia - Protein's conformational space studies using perturbation-relaxation MD methods.

Software demonstrations.

IV. (Continuation)
Models for protein folding and function
Chairperson: C. Sander & O. Tapia

K. A. Dill - Simple lattice models of protein folding.

M. Levitt - Computer simulation of protein folding and unfolding.

M. Karplus - Protein folding simulations: From the native to the denatured state and back again.

Software demonstrations.

INTRODUCTION

F. X. Avilés, M. Billeter and E. Querol

INTRODUCTION

The challenge of understanding the structure and function of proteins lies on the one hand in the high complexity of these molecules and on the other hand in the vast information in terms of sequences and three-dimensional structures accumulated over the past years. The obvious tools to handle this vast information are computers. They allow , among others, approaches such as the assembly and use of data bases for prediction purposes, the construction and testing of theoretical models, the simulation of the short-time dynamical behaviour of proteins , and the design of new proteins.

Computational molecular biology is already a discipline in its own right. It is now part of the daily routine of many biologists, but also a useful tool to explore the frontiers of science in the world of biomacromolecules, such as proteins and ligands. Biological sciences are now posing questions that not only require computational solutions, but which also provide problems of fundamental interest to computer scientists, creating the need for an interdisciplinary work to develop data adquisition systems, adequate algorithms for data comparision, for prediction of the conformation and function of proteins, for ligand and drug design...etc.

The present workshop provided an opportunity for interdisciplinary researchers to discuss on the frontiers of biocomputing of proteins, covering various areas :

-Data bases comprising both sequence and three-dimensional data were presented. The classification of proteins and the derivation of folding determinants common to certain protein families rely on these data bases, and still provide the starting point for many structure prediction concepts. On the other hand, the inverse question addressing the correctness of a three-dimensional structure for a given sequence can today in most cases be answered with good reliability.

-Theoretical models like the folding of a point chain in a rectangular grid make us reconsider the origin and the role of central folding elements such as regular secondary structures. Another major problem is the localization of the "correct" structure, or with other words of a good energy minimum, even when a start structure with the native backbone folding topology is provided. Current methods such as Molecular Dynamics or Monte Carlo algorithms exhibit a rather small radius of convergence. Newer methodologies that were developed and successfully applied in other areas, an example being the mathematical theory of fuzzy logic, might well become of considerable interest but are still in a very early stage.

-The potentiality and use of computational approaches for the structure-based drug design was another topic showing a great progression at both the theoretical and applied level.

-The presentation of studies on individual systems showed, in spite of the above mentioned difficulties, very encouraging successes

It was exciting to bring together such a diverse set of issues and participants whose active involvement made the level of discussions both animated and very demanding. On behalf of all the organizers and participants we wish to acknowledge

the opportunity for such a stimulating scientific event provided by the Instituto Juan March . The efficient organization and the generous support of its members resulted in a highly productive workshop. We also wish to acknowledge the collaboration of the firms Silicon Graphics, Biosym and Tripos which provided hardware and software and contributed to the organization of computer-based demonstrations during the meeting.

Madrid, November 1994

Francesc X. Avilés
Martin Billeter
Enrique Querol

I. Prospects in Computational Molecular Biology

Protein structure: 1D-2D-3D

Chris Sander, EMBL, D-69012 Heidelberg, Europe

Protein folding is a classical problem of molecular biology. Recent progress has been made in three areas.

Structure prediction in 1D

Prediction of secondary structure from sequence alone has surpassed the 70% three-state accuracy level using a neural network algorithm with multiple related sequences as input [1,2].

Structure comparison in 2D

Representing three-dimensional structures as two-dimensional contact maps, a new algorithm for the identification of structural similarities in the database of known structures has been developed. An all-against-all comparison of all known protein structures leads to the discovery of unexpected stuctural relationships and insights into the spectrum of naturally occuring protein folds [3,4].

Prediction of contacts in 3D

Analyzing correlated mutational behavior in multiple sequence alignments, an estimate can be made of which residues are in contact in the three-dimensional structure. For the most correlated residue pairs predicted to be in contact, the prediction accuracy ranges from 37% to 68% and the improvement ratio relative to a random prediction from 1.4 to 5.1. Predicted contact maps may be useful as input for the calculation of protein tertiary structure [5].

[1] B. Rost, C. Sander, *Combining evolutionary information and neural networks to predict protein secondary structure*
Proteins **19**, 55-72 (1994)

[2] B. Rost, C. Sander, *Prediction of protein secondary structure at better than 70% accuracy*
J.Mol.Biol. **232**, 584-599 (1993).

[3] L.Holm, C.Sander,*Protein structure comparison by alignment of distance matrices*
J.Mol.Biol. **233**, 123-138 (1993).

[4] L.Holm, C.Sander,*The FSSP data base of structurally aligned protein fold families*, Nucleic Acids Research **21**, 3105-3109. (1993)

[5] U. Goebel, C. Sander, R. Schneider, A. Valencia, *Correlated mutations and residue contacts in proteins* , Proteins **18**, 309-317 (1994)

3D PROFILES FOR ASSESSING, DESIGNING, AND REFINING PROTEINS

David Eisenberg, James U. Bowie, Roland Lüthy, Laura Wesson, Scott Le Grand, Kam Zhang, Arne Elofsson, Matthias Wilmanns

UCLA-DOE Lab of Structural Biology and Molecular Medicine, Molecular Biology Institute, and Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1570. Phone 310 825-3754. FAX 310 206-3914. E-mail: DAVID@UCLAUE.MBI.UCLA.EDU

The compatibility of an amino acid sequence with a given protein structure can be measured with a 3D profile. The 3D profile assesses the compatibility of a sequence with a structure by summing scores for the compatibility of each amino acid residue with its environment in the 3D structure. The 3D profile is a matrix which gives the compatibility ('1D-3D score') for each of the 20 amino acids in the environments of all residues of a structure. The profile permits searching of the sequence database for sequences compatible with the structure. Alternatively, a given amino acid sequence can be compared to a library of 3D profiles.

3D profiles have found two main applications. The first is the assignment of amino acid sequences to known 3D structures. The second is assessment of the correctness of protein models, whether determined by X-ray, NMR, or computational methods. The 3D profile of a correct protein model gives a high score for compatibility with its sequence (examples include coordinate sets in the Protein Data Bank), whereas the profile of an incorrect model gives a low score for compatibility with its own sequence (examples include mistraced proteins, as well as deliberately misfolded proteins).

Two related applications of 3D profiles are the assessment of designs for proteins, and the assessment of the state of refinement of a protein model determined by x-ray crystallography. Examples of these applications will be given.

Profiles can also aid in directing protein folding calculations by an evolutionary algorithm, starting with only the amino acid sequence. In this approach, the sequence to be folded is broken into segments. Likely conformations for each segment are selected by the compatibility of the sequence with profile fragments. By linking selected conformational fragments, hundreds of trial structures are generated, each having the same length and sequence as the protein to be folded. These starting trial structures are then improved by an evolutionary algorithm in which the structures are altered by changes in the dihedral angles and by recombining segments of the structures. Selection pressure for improving the structures is provided by a fitness function that is designed to guide the evolu-

tionary process smoothly toward the true structure. This fitness function includes the profile score. We find that by evolution of only 400 structures for fewer than 1400 generations, the overall fold of some small helical proteins can be computed from the sequence, with deviations from observed structures of 2.5-3.5 Å for alpha carbon atoms.

Computer simulation of proteins, prospects and limitations

Wilfred F. van Gunsteren
Laboratory of Physical Chemistry
Swiss Federal Institute of Technology Zürich
ETH Zentrum, CH-8092 Zürich
Switzerland

In the past decade computef simulation has become a useful tool to investigate the structural, energetic and dynamical properties of biomolecules at the atomic level. Its application involves three basic choices:

(1) The degrees of freedom of the molecular model should be chosen such that the molecular properties one is interested in are adequately modelled.

(2) An atomic interaction function describing the interaction between the chosen degrees of freedom, which should be of sufficient accuracy to yield molecular properties to the desired degree of accuracy.

(3) Equations of motion or an algorithm to generate an ensemble of molecular configurations from which the molecular properties can be computed by taking trajectory or ensemble averages. The simulation or sampling period should be taken sufficiently long to obtain a converged equilibrium ensemble.

The various choices and approximations involved in practical simulation studies will be discussed and illustrated with practical examples.

References:

- W.F. van Gunsteren
Molecular dynamics and stochastic dynamics simulation: A primer
In: "Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications", Vol. 2, W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson eds., Escom Science Publishers, Leiden, The Netherlands, (1993), pp. 3-36

- W.F. van Gunsteren, T.C. Beutler, F. Fraternali, P.M. King, A.E. Mark and P.E. Smith
Computation of free energy in practice: choice of approximations and accuracy limiting factors
In: "Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications", Vol. 2, W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson eds., Escom Science Publishers, Leiden, The Netherlands, (1993), pp. 315-348

- W.F. van Gunsteren
Molecular dynamics studies of proteins
Current Opinion in Structural Biology 3 (1993) 277-281

- W.F. van Gunsteren, R.M. Brunne, P. Gros, R.C. van Schaik, C.A. Schiffer and A.E. Torda
Accounting for Molecular Mobility in Structure Determination Based on Nuclear Magnetic Resonance Spectroscopic and X-Ray Diffraction Data
in: "Methods in Enzymology: Nuclear Magnetic Resonance", Vol. 239, T.L. James, N.J. Oppenheimer eds., Academic Press, New York, (1994), pp. 619-654

- W.F. van Gunsteren, F.J. Luque, D. Timms and A.E. Torda
MOLECULAR MECHANICS IN BIOLOGY: From Structure to Function, Taking Account of Solvation
Ann. Rev. Biophys. Biomol. Structure 23 (1994) 847-863

II. Data bases, prediction and modelling

Rule-based approaches to the prediction and modelling of proteins.

Tom Blundell, ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E7HX.

Analogies between distantly related, possibly convergently-evolved super-families of proteins, can be as useful in the prediction of protein structure as the well defined homologies, long exploited in comparative modelling. We require an automated procedure that uses the known structures and the rules obtained from them - the knowledge base. We first formulate rules, not only by analysing individual protein structures, but also by comparing sequences and 3-D structures. For this we need methods not only for comparing many sequences and three-dimensional structures in order to establish equivalences within a family of proteins. The rules are then derived by careful, computerised analysis of the compared structures. This is the learning stage. These rules can then be applied in two important ways. First, they are used to define all those sequences that can adopt each experimentally defined fold; this can be considered as a projection from 3-D structure onto 1-D sequence and is the inverse folding problem. It is usually achieved through construction of a template, which is a generalized protein sequence summarizing knowledge about the family fold and presented in a form suitable for comparison with the sequence of the "unknown". Secondly, when the alignment of sequence with the template has been achieved, the rules are used together in an extrapolation of features of the known structures to the sequence of the protein to be modelled. The three-dimensional coordinates of the complete model are derived so that they satisfy the spatial restraints implied in this process of extrapolation as well the general rules of protein structure. I shall describe computational approaches to these different stages of protein structure prediction that have been developed at Birkbeck.

Protein Folds and their Determinants

Janet Thornton, David Jones, Malcolm MacArthur, Christine Orengo & Mark Swindells
Biochemistry and Molecular Biology Dept.
University College
Gower Street

There are now over 3000 polypeptide chains in the Brookhaven Protein Structure Databank. These structures provide an invaluable resource, from which we can learn more about the rules which govern the formation of three dimensional structure. In this presentation two different aspects of the relationship between structure and sequence will be presented. Firstly I shall report our recent studies of ϕ,ψ propensities derived from the coil regions of protein structures. These are intrinsic preferences which can only reflect the local interactions between the side chains and the adjacent peptide units. We find that, with a few exceptions, there is a good correlation between these preferences and secondary structure propensities. Then, moving from secondary structure to protein topology, a review of the protein domain folds, as determined by structure comparison methods will be given, which has highlighted the very uneven distribution of naturally occuring folds. We find that a few folds completely dominate the protein structure database, comprising over 50% of the non-homologous proteins. This has implications for protein folding and fold recognition which will be discussed.

Information processing and the need of biological data bases: An overview and a case study on the 3D density data base

J.M.Carazo, Centro Nacional de Biotecnología-CSIC, Univ.Autónoma, 28049 Madrid. Tf.: 341 585 4543. Fax: 341 585 4506, e-mail: carazo@samba.cnb.uam.es

The topic of this lecture is centered on the need to organize the increasing flow of data that is arising in modern biology. Indeed, this new information is having a crucial impact in many aspects of information processing. In the first place, it pushes the development of efficient ways to organize the information in the form of data bases. In the second place, the availability of well organized and vast amount of data immediately highlights the need to design tools that made possible the understanding and interrelation of this information. A third aspect, that naturally comes into play as a consequence of the globalization of most modern human activities, is the need to make accessible this body of information to the whole of the research community world wide.

The available list of data bases in Biology is certainly very large, so large that, in fact, it constitutes a data base in itself!. However, it is true that not all of them have the same frequency of usage among the molecular biology community, and some of the most widely used data bases will be commented in this presentation. On the topic of tools to analyze the data, I will make a special emphasis on the need to interrelate -or "link"- many pieces of information from different data bases in order to obtain a global view of a given problem. Some examples on how textual link can indeed "discover" new information from information already available, though disperse, will be presented. However, the need to go much beyond textual links will drive the presentation. The very actual need to provide information world wide will be specifically addressed, providing a short exposition of the actual and near future possibilities of this new topic.

Throughout the presentation I will be using as a case study the data base that we work on, that is formed by 3D densities distributions obtained by 3D electron microscopy. In essence, the entries to this data base are volume data obtained by tomographically reconstructing in three dimensions a given macromolecular aggregate. The typical resolution range is around 2-3 nm, although some examples at resolutions below 1 nm will be presented. Textual links to sequence information, bibliography or sets of atomic coordinates will be introduced at this stage. The need to develop more elaborated links will be presented through a "proof of concept" example on a possible link between this data base and PDB. A step by step session of a hypertext-based query to this new data base will also be presented.

JOHN DEVEREUX

Genetics Computer Group, University Research Park
575 Science Drive
Madison, WI. 53711-1060 (USA)

Advances in Amino Acid Substitution Matrices.

Amino acid substitution frequencies were measured by Dayhoff. These measurements were the basis of the widely used PAM series of amino acid substitution matrices which have be used to create most of the optimal alignments with which the literature is now filled. The recognition that these PAM substitution matrices were really log-odds values and that, in this form, a sum of such scores could be used to state the probably of observing a gapless alignment was buried in the work of Dayhoff from whence it was unearthed by Karlin and Altschul.

Specifically, Karlin and Altschul predicted that, for any PAM matrix, you could predict the maximum score of a short gapless alignment that would arise by chance from a comparison of any given size. Armed with this information you could set a cutoff score for sequence database searches that would protect you from being showered with alignments that have no way to be distinguished from chance. While Karlin-Altschul mathematics, which forms the foundation of BLAST, is firmly rooted in the prediction of the target frequencies for amino acid substitutions in proteins diverged by a known amount, recent work by Henikoff and Henikoff has generated a substitution matrix, which works better than the PAM matrices in most protein sequence database searches. Unfortunately the Henikoff matrices do not fit the Karlin-Altschul theory as neatly as one could have hoped.

Structural and functional information contained in protein families: Engineering of specificity

Alfonso Valencia.

European Molecular Biology Laboratory, Heidelberg and Centro Nacional de Biotecnologia, CSIC, Madrid. Internet: valencia@EMBL-Heidelberg.DE

Evolution has been playing with protein sequences for a long time. Selective pressure maintains protein function, which in turn requires maintenance of the specific three-dimensional structure consistent with that function. Accordingly, conservation and mutation patterns observed in multiple sequence alignments are evidence of functional or structural constraints plus mutational drift.

The increasing number of protein families for which many homologous sequences are known affords a new opportunity to exploit evolutionary information.

I will present here two parallel approaches: prediction of inter-residue contact and localization of sites of specificity. In both cases only multiple sequence alignments are used as input data.

First we were interested in extracting information about 3D contacts from a family of sequences (1). The key assumption here is that residues in physical contact show correlated mutational behaviour: if one residue mutates, its contact partners also tend to mutate. In addition, we assume that this effect is detectable at the level of pair correlations. Higher order clusters are then built up from mutually correlated pairs.

The method to analyse correlations in mutational behaviour between different positions is based on defining an exchange matrix at each sequence position in a multiple alignment and then calculating a correlation coefficient between the exchange matrices at any two positions. The map of position-position correlations is interpreted as a predicted residue-residue contact map and is then compared with the contact map from experimentally observed structures. The extent of agreement between the two indicates of the validity of the underlying assumptions and of the

usefulness of the predicted contact maps for prediction of protein three-dimensional structures. For the most strongly correlated residue pairs predicted to be in contact, the prediction accuracy ranges from 37% to 68% and the improvement ratio relative to a random prediction from 1.4 to 5.1. These predictions are much better than other possible ones, i.e. simple predictions of contact maps with pair-residue preferences derived from the known 3D structures (2).

In second place we ask which residues are important for specifying the divergence between different sub-families (tree determinant residues). Many protein families contain more than one sub-family with different functionality or specificity, i.e. protein kinases are divided in serin-threonin and tyrosin kinases, and serin-threonin kinases at the same time can be subdivided in cAMP dependent kinases, protein kinases C, dcd2 like, etc. This divergence between sub-families is driven by functional constrains and subject of strong selective pressure.

We analysed simultaneously the sub-family, sequence and position specific information with a Principal Component method. Gerog Casari is presenting in this meeting a detailed description of the method. The results obtained in well-characterised systems as p21-ras or sh2 domains are very encouraging (3).

A further development in this line is the implementation of the analysis tool for the classification of protein sequences and detection of tree-determinant residues using a Kohonen auto-organising neural network (4). This classification device allows to cluster protein sequences in groups and further determination of tree determinant residues.

The success of the predictions has been experimentally tested. After detailed prediction and modelling of the regions responsible for differences between groups of sequences it was possible to exchange the function between two proteins of the rab-subfamily (5).

(1) Gobel, U., Sander, C., Schneider, R. & Valencia, A. (1994) Proteins, 18: 309-317.
(2) Tamames, J., Sander, C. & Valencia, A. (1994) in preparation.
(3) Casari, G., Sander, C. & Valencia, A. (1994) Submitted.
(4) Andrade, M. A., Casari, G., Sander, C. & Valencia, A. (1994) in preparation.
(5) Stenmark, H., Valencia, A., Martinez, O., Ullrich, O., Goud, B. & Zerial, M. (1994) EMBO J., 13: 575- 583.

A Hierarchical Simulation Method of Protein Folding

Jeffrey Skolnick and Andrzej Kolinski

Dept. Molecular Biology. The Scripps Research Institute. USA.

Antonio Rey

Dept. Química Física I. Universidad Complutense de Madrid. Spain.

The protein folding problem refers to the prediction of the three-dimensional structure of a globular protein from its primary sequence of amino acids. We propose a method to approach this problem by the use of a set of computer simulation models which progressively represent the geometry of a globular protein from a coarser to a finer resolution [1]. The set of interactions responsible of stabilizing a unique folded conformation is described as well. Our aim is to set up a computer model which, starting from the amino acid sequence alone, generates a folded native structure. By using this model to simulate the time evolution of the system, one could in principle get information about the folding pathways as well, therefore collecting some insight into possible folding intermediates.

The geometrical details of the model are based on the projection of the protein backbone on an underlying cubic lattice. Different lattice grids are employed [2]. At the beginning of the procedure, a coarse grid allows for an approximate geometrical representation of the protein backbone, but the accompanying capital reduction in the number of degrees of freedom makes it possible to very efficiently simulate the wide conformational space available for the extended states of an unfolded protein, the starting point of our conformational search. When compact conformations are obtained, they are projected onto a finer grid lattice, which permits a better geometrical description of the protein geometry. This is more adequately suited to describe the packing effects and other details important to distinguish between compact conformations as those appearing in a molten globule and a real folded conformation.

The dynamics of the system is simulated by the use of a very efficient Monte Carlo algorithm [2], which randomly explores the protein model conformations through a set of small displacements of the units representing the amino acids. These Monte Carlo movements are carefully balanced to permit motions of single units, a small number of them, or even a full element of secondary structure. Thus, a kinetic folding mechanism is not *a priori*

assumed.

The conformational search performed by the Monte Carlo procedure is driven by the energy function defined for the protein model. We have used several mean field components to the energy, predominantly based on a statistical analysis of known protein structures. The main contributions to this energy function are: (i) a hydrogen bond network for the backbone, based on the geometrical arrangements of α–carbons in the model. (ii) An angular correlation between side-chain orientations with respect to the backbone, for amino acids which are close along the protein sequence. (iii) A pairwise interaction between amino acid side groups. And (iv), a contact map template interaction which tends to favour cluster of contacts in the protein core as those encountered in real proteins. Both the hydrogen bond network and the template contribution create a set of highly cooperative interactions which can favour an "ordered collapse" of the protein chain to a compact conformation with a simultaneous development of the elements of secondary structure. The template part of the potential can also yield the specific packing of the protein interior which can differentiate a folded globular protein interior from a molten globule intermediate.

The model is not accurate enough yet to provide a unique structure for any sequence. However, examples are presented of its possibilities in determining adequate secondary structure elements and even correct tertiary folds corresponding to sequences of real proteins with simple topologies [3].

References

[1] J. Skolnick, A. Kolinski, C.L. Brooks III, A. Godzik, and A. Rey, *Current Biology* **3**, 414 (1993).

[2] A. Kolinski and J. Skolnick, *Proteins* **18**,338 (1994).

[3] A. Kolinski and J. Skolnick, *Proteins* **18**,353 (1994).

EXTRACTING INFORMATION ON THE PROTEIN 3D STRUCTURE AND FOLDING PATHWAY FROM THE AMINO-ACID SEQUENCE

Shoshana J. Wodak, Marianne J. Rooman, Jean Pierre Kocher, and René Wintjens

Unité de Conformation de Macromolécules Biologiques, Université Libre de Bruxelles, CP160/16, av. Paul Héger, B1050 Brussels (Belgium)

There has been a renewed interest lately in deriving effective potentials, based on reduced descriptions of the protein conformation, from the growing number of protein structures. These potentials are most commonly obtained from statistical relations between the amino-acid sequence and structural features. They are viewed by many as a promising alternative to classical empirical potentials, because they yield a much smoother energy hypersurface, which is bound to increase the efficiency of conformational search procedures, and seem capable of recognizing the native fold of a given sequence from a number of incorrect alternatives. Here we illustrate two applications of such potentials to the prediction of local structure in proteins.

In one application, effective potential terms, which incorporate only local interactions between residues along the chain, are used to predict the conformation of short peptides and of regions along the polypeptide chain that adopt a well defined local structure in absence of interactions with the rest of the protein. A survey of such regions in 69 non-homologous proteins from the Brookhaven Databank, and in 12 families and subfamilies of homologous proteins, indicates that they correspond primarily to segments of secondary structure heavily involved in interactions with the rest of the protein. This suggests that these regions could act as nuclei around which other parts of the structure would assemble, which in many cases is supported by experimental data on early folding intermediates.

In another application, we present a systematic evaluation of the ability of effective potentials which incorporate different contributions, including, residue-residue interactions, and solvation terms, to identify short sequence segments (10-30 residues long) adopting specific turn motifs in proteins. Families of such turn motifs are first identified in the dataset of known protein structures by a fully automatic classification procedure. Several types of fingerprints are then derived for each family, and all the sequence segments in the dataset are successively mounted onto the fingerprints, with the energy of each mount being evaluated using the effective potentials. The results show that the success with which sequence segment recognize native, or compatible, turn motifs remains rather limited to a level which is much below the native recognition performance generally achieved for full proteins. The fingerprints and their associated potentials are nevertheless found to be quite effective in generating subsets of solutions which are significantly enriched for the correct sequence-structure combinations. This may have very useful applications in protein folding simulations and in homology modelling.

Knowledge based force fields in the analysis, recognition and computation of protein folds

Manfred J.Sippl

A major goal in protein structure theory is the development of energy functions which can be applied in the analysis, recognition and ab initio calculation of biological macromolecules. In particular the unsolved problem of predicting the three dimensional structure of proteins solely from the information contained in amino acid sequences is in the forefront of biophysical chemistry. Remarkable resources have been devoted to tackle this problem and the field has seen some progress in recent years. Among the new approaches are knowledge based techniques. Their main goal is the extraction of energy functions from experimental data on molecular structures and their application in structure determination and engineering of biological molecules.

In particular the set of protein structures solved over the last decades contains a vast amount of information on molecular interactions which stabilize native protein folds in solution. Basic principles of statistical mechanics enable the extraction of this information in terms of energy functions. The resulting knowledge based force field consists of a set of mean force potentials describing molecular interactions with respect to an average protein-solvent environment.

We discuss the basic steps in the compilation of mean force potentials from a data base of protein structures and demonstrate their predictive power at the present state of development. The applications presented cover the recognition of errors in protein structures, the recognition of sequences which adopt a given fold, and the calculation of differences in stability ($\Delta\Delta G$-values) of wild type and mutant proteins. We finally glance at current activities and future developments.

References:

Sippl,M.J., Boltzmann's principle, knowledge based mean fields and protein folding. Journal of Computer Aided Molecular Design, 7 (1993), 473-501.

Sippl,M.J., Recognition of errors in three dimensional structures of proteins. Proteins, 17 (1993), 355-362

Sippl,M.J., Calculation of conformational ensembles from potentials of mean force. Journal of Molecular Biology, 213 (1990), 859-883

Contacts to China:

I am collaborating with Dr.Feng Zu Kang (Academia Sinica, Shanghai Institute of Biochemistry) on several aspects of the three dimensional structure of proteins. Dr.Feng Zu Kang is a post doctoral fellow (Austrian Science Foundation, Liese Meitner Stipendium) in my group.

III. Molecular recognition and drug design

Role of Homology Modeling in the Design of Novel Pharmaceutical Agents

Jonathan Greer
Abbott Laboratories

Abstract

One of the major challenges of structural biology today is the use of three-dimensional structure of target protein molecules and their complexes with ligands to aid in the design of novel inhibitors, agonists, or antagonists as new pharmaceutical agents.[1] Crucial to this process is the timely solution of the experimental structures of the protein/ligand complexes by either crystallographic or nuclear magnetic resonance (NMR) methods. Both of these methods require large quantities of the protein which are usually not readily available; they need to be cloned, expressed, and purified using molecular biology and biochemistry techniques. In the interim, a starting structure can frequently be produced using homology modeling.

A second important use for homology modeling is emerging from the success of the human genome project and related mass sequencing efforts. Is it anticipated that the full human genome, including perhaps 100,000 proteins, will be sequenced in the next ten to fifteen years. Determination of the three-dimensional structures of these proteins will take another century, probably, if they must be obtained using current experimental methodologies. Homology modeling holds the potential of providing approximate structural information for many of these proteins very rapidly to aid in understanding the function of these molecules and in evaluating their potential as therapeutic targets.

Homology modeling can performed on any protein which is homologous in sequence, and hence in three-dimensional structure, to a least one other protein who structure has been determined experimentally. The closer the sequence homology, the closer the structural homology, and the more accurate the model is likely to be. Methods to construct such models will be described.[2,3] The models produced by these methods are only tentative; the methods do permit confidence limits to be assigned to the various parts of the structure.

Homology model structures will be illustrated by examples from the serine protease family. The purpose of the large number of proteins in this family, all sharing the same catalytic site and mechanism, is to permit high specificity of a particular protease for its physiological substrate. Using the interaction of the blood clotting factor Xa with its substrate prothrombin, the basis of the exquisite specificity of these enzymes was examined.[4] One member of this family, prostate specific antigen, is used for diagnostically to detect prostate cancer. Homology modeling of this protein has been employed to optimize the diagnostic use of this protein. Another of the proteins modeled is urokinase, a plasminogen activator involved in fibrin clot lysis and cell-cell interactions. A three-dimensional model of urokinase was developed and used to help understand the structure-activity of a new family of specific inhibitors, the phenylguanidines.[5] The model was also valuable in rationalizing the abnormal enzymatic properties of a group of mutated urokinases.[6] The nature of specific protein-protein interactions and aggregation was studied by docking a model of haptoglobin, a non-enzymatic serine protease homolog, onto a hemoglobin molecule to propose a three-dimensional structure for the extremely stable complex of these two proteins.

With a large range of uses for homology modeled proteins and ever increasing role anticipated with the advent of genomic sequencing, it becomes important to analyze how accurate the models actually are. Then, as a crucial follow-up question, can such an analysis suggest strategies for improving the accuracy of such models? A number of the models that have been proposed over the past decade and a half have since had their actual structures determined experimentally. These models include two serine proteases and three aspartic proteinases. We have begun a detailed analysis of the differences between the modeled and experimental structures to aid in improving protein structure construction using homology. The model structure of blood clotting factor Xa was recently used by Tulinsky and coworkers to solve the crystal structure of factor Xa.[7] The model structure is being compared to the refined crystallographic structure. Similar studies are in progress for the model of the blood hypertension protein renin,[8] an aspartic proteinase. Its structure is being compared to experimental structures that have been determined independently in three different laboratories.[9-11] Some of the lessons learned from these comparisons will be described.

References

(1) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **1994**, *37*, 1037-1054.

(2) Greer, J. Comparative model-building of the mammalian serine proteases. *J Mol Biol* **1981**, *153*, 1027-1042.

(3) Greer, J. Comparative Modeling Methods: Application to the Family of the Mammalian Serine Proteases. *Proteins* **1990**, *7*, 317-334.

(4) Greer, J. Model of a specific interaction. Salt-bridges form between prothrombin and its activating enzyme blood clotting factor Xa. *J Mol Biol* **1981**, *153*, 1043-53.

(5) Yang, H.; Henkin, J.; Kim, K. H.; Greer, J. Selective inhibition of urokinase by substituted phenylguanidines: QSAR analyses. *J. Med. Chem.* **1990**, *33*, 2956-2961.

(6) Davidow, L. S.; Dumais, D. R.; Smyth, A. P.; Greer, J.; Moir, D. T. Mutations affecting the activity of urokinase-type plasminogen activator. *Protein Engineering* **1991**, *4*, 923-928.

(7) Padmanabhan, K.; Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kisiel, W. Structure of human des(1-45) factor Xa at 2.2 A resolution. *J Mol Biol* **1993**, *232*, 947-66.

(8) Hutchins, C.; Greer, J. Comparative modeling of proteins in the design of novel renin inhibitors. *Crit. Revs. Biochem. Molec. Biol.* **1991**, *26*, 77-127.

(9) Sielecki, A. R.; Hayakawa, K.; Fujinaga, M.; Murphy, M. E.; Fraser, M.; Muir, A. K.; Carilli, C. T.; Lewicki, J. A.; Baxter, J. D.; James, M. N. Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5 A resolution. *Science* **1989**, *243*, 1346-51.

(10) Dhanaraj, V.; Dealwis, C. G.; Frazao, C.; Badasso, M.; Sibanda, B. L.; Tickle, I. J.; Cooper, J. B.; Driessen, H. P.; Newman, M.; Aguilar, C.; et, a. l. X-ray analyses of peptide-inhibitor complexes define the structural basis of specificity for human and mouse renins. *Nature* **1992**, *357*, 466-72.

(11) Rahuel, J.; Priestle, J. P.; Grutter, M. G. The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J Struct Biol* **1991**, *107*, 227-36.

MOLECULAR DYNAMICS AND FREE ENERGY PERTURBATION STUDIES OF THE INTERACTION BETWEEN DRUGS AND PROTEINS

Modesto Orozco

Departament de Bioquímica i Fisiologia. Facultat de Química. Universitat de Barcelona. Martí i Franquès 1. Barcelona 08028.

Molecular Dynamics (MD) is one of the most widely used and powerful techniques to study the behavior of biological macromolecules in physiological-like environments at the atomic level. Furthermore, the combination of MD and statistical mechanics algorithms facilitates the determination of time-averaged properties of macromolecular systems, including the calculation of the free energy differences associated with conformational changes and molecular interactions. Among the different techniques based on statistical mechanics free energy perturbation (FEP) is one of the most popular. This technique determines the free energy difference between two similar stationary states from the exponential average of the energy difference between both states. The Boltzman sampling necessary for the averaging is obtained from MD or Monte Carlo (MC) simulations.

The goodness of the results derived from MD/FEP calculations depends mainly on: i) the similarity between the real and simulated systems, ii) the ability of the classical force-field to represent the molecular Hamiltonian, iii) the ability of the MD simulation to sample properly the conformational space accessible to the system. It must be emphasized that the reliability of the MD/FEP simulations will be strongly determined by the goodness of the simulation protocol, and on the nature of the problem considered.

The use of MD/FEP simulations in the study of protein-drug interactions is illustrated by three different simulations. The first simulation illustrates the use of MD/FEP methods to obtain approximated models of drug-protein interactions. The second simulation allows us to predict differences in binding free energy between related substrates. The last simulation shows how the MD/FEP methods can be used to gain qualitative insight into mechanistic aspects of protein action. The three simulations presents a series of technical problems which defines the range of uncertainties of the results, and accordingly the validity of the conclusions that can be reached from the calculations.

Toward a more rational design of enzyme inhibitors: phospholipase A_2 as a test case.

Angel. R. Ortiz, M. Teresa Pisabarro, and Federico Gago
Departamento de Fisiología y Farmacología, Universidad de Alcalá de Henares, Madrid

The results of a project aimed at designing specific inhibitors for human synovial fluid phospholipase A_2 (HSF-PLA_2) enzyme using structure-based computational methods will be presented. This enzyme, which is found in high concentrations in the synovial fluid of patients with rheumatoid arthritis, catalyzes the hydrolysis of aggregated phospholipids at the *sn-2* position. Release of arachidonic acid from this position is the rate-limiting step in the biosynthesis of several eicosanoids involved in inflammation, and it is thought that inhibitors of this secretory enzyme, as well as inhibitors of a different high molecular weight cytosolic PLA_2, may become a new class of interesting pharmacological agents (1).

In order to design effective inhibitors, it was first necessary to understand the basic features of interfacial catalysis in structural terms as differential interactions at the phospholipid-enzyme interface can control the potencies of PLA_2 inhibitors. By modelling the interactions of several PLA_2s with different phospholipid head groups, we proposed the existence of an interfacial binding site in a variable loop region of PLA_2 enzymes which could account for the distinct substrate preferences of some members of this family (2).

The next step was to try to understand the mechanism of action of known PLA_2 inhibitors at the molecular level. Manoalide (MLD), a natural product from a marine sponge, is a potent inhibitor of PLA_2s from different sources and is thought to exert this effect by interfering with the process of interfacial activation. We proposed an active conformation for MLD and also identified a putative binding site for MLD on bee venom PLA_2 separate from the active site, as the experimental evidence suggested. This binding site appears to overlap the interfacial binding site previously described. From the standpoint of ligand design, the main conclusion from this piece of work was that the specific MLD binding site found in bee venom PLA_2 is not present in other members of the family. Thus, this enzyme cannot be considered a good model system on which to design and test novel anti-inflammatory MLD-like compounds directed against proinflammatory PLA_2s (3).

We then turned to the active site of HSF-PLA_2. When the three dimensional structure of this enzyme was known both in its native form and in a complex with the transition state analogue (TSA) L-1-*O*-octyl-2-heptyl-phosphonyl-*sn*-glycero-3-phospho-ethanolamine (4), we attempted to modify this particular inhibitor in order to improve binding. Different probes within the GRID program (5) were used to generate contour maps throughout the enzyme which could give us an indication of regions of space favouring particular functional groups and suggesting possible extensions of the inhibitor that would make use of additional binding sites. A hydrophobic pocket, delimited by a seven-residue C-terminal extension of the enzyme, was detected that could be exploited to achieve increased binding affinity. Additionally, this region could be used as a target for incorporating pharmacological selectivity into the inhibitors

since it does not exist in the pancreatic enzymes. The LUDI program (6) was employed to identify and orientate suitable chemical fragments in this cavity, and different compounds were synthesized and tested. One of this (LM-1228) was shown to be a potent HSF-PLA_2 inhibitor indeed possessing the predicted differential activity toward HSF and pancreatic PLA_2 enzymes (7).

A series of TSAs generated using the methodology outlined above constitutes a valuable source of information regarding the requirements for both binding affinity and biological activity of a set of compounds. Such information can in principle be extracted by means of Quantitative Structure-Activity Relationships (QSAR). Nevertheless, current QSAR approaches are hampered by limited use of structural information about the ligand-receptor complexes in the derivation of correlations. To overcome this limitation, we have developed a new approach that dissects calculated ligand-receptor interaction energies into a large number of terms (independent variables) and subjects them to statistical analysis in order to highlight the subset of energy contributions which are responsible for most of the variance in binding free energy (response). The method, which we term *Com*parative *Bin*ding *E*nergy (*COMBINE*) analysis, attempts to separate the "energetic signals" from the "background noise", and has been first applied to a series of 26 HSF-PLA_2 inhibitors (8). Once the relevant components are selected and adequately scaled, a regression equation is obtained in which activity is correlated with the interaction energies between parts of the ligand and key residues of the protein. The resulting model is highly predictive and compares favourably with standard 3D-QSAR techniques like CoMFA. The implications of this new method for rational drug design will be discussed.

References

1. Glaser, K. B., Mobilio, D., Chang, J. Y., Senko, N. *TiPS* **14**, 92-98 (1993).
2. Ortiz, A. R., Pisabarro, M. T., Gallego, J., Gago, F. *Biochemistry* **31**, 2887-2896 (1992).
3. Ortiz, A. R., Pisabarro, M. T., Gago, F. *J. Med. Chem.* **36**, 1866-1879 (1993).
4. Scott, D. L., White, S. P., Browning, J. L., Rosa, J. J., Gelb, M. H., Sigler, P. B. *Science* **254**, 1007-1010 (1991).
5. (a) Goodford, P. J. *J. Med. Chem.* **28**, 849-857 (1985). (b) GRIN, GRID and GRAB programmes, Molecular Discovery Ltd., West Way House, Elms Parade, Oxford OX2 9LL, England.
6. (a) Böhm, H.-J. *J. Comput. Aided Mol. Design* **6**, 61-78 (1992). (b) Böhm, H.-J. *J. Comput. Aided Mol. Design* **6**, 593-606 (1992). (c) Biosym Technologies, 9685 Scranton Road, San Diego, CA 92121-2777.
7. Pisabarro, M. T., Ortiz, A. R., Palomer, A., Cabré, F., García, L., Wade, R., Gago, F., Mauleón, D., Carganico, G. *J. Med. Chem.* **37**, 337-341 (1994).
8. Ortiz, A. R., Pisabarro, M. T., Gago, F., Wade, R. (submitted).

Structure-Based Drug Design of VX-478, a Potent and Orally-Bioavailable Inhibitor of the Human Immunodeficiency Virus Type 1 (HIV-1) Protease.

M. A. Navia, E. E. Kim, C. T. Baker, M. D. Dwyer, D. D. Deininger, M. A. Murcko, B. G. Rao, R. D. Tung
Vertex Pharmaceuticals Incorporated; 40 Allston Street; Cambridge, Massachusetts 02139-4211; U. S. A.

Significant advances in the technologies of structure determination and molecular modeling have fostered the accelerated development of structure-based drug design as an applied science. Structural information - to be exploited for the enhancement of *in vitro* potency - can now be anticipated with some predictability for newly identified therapeutic target enzymes and their inhibitor complexes. Initial steps have also been taken to apply the paradigm of structure-based drug design to the broader problems of *in vivo* efficacy, including bioavailability, toxicity, and clearance. Examples of this comprehensive approach will be discussed, with particular emphasis on the inhibition of HIV-1 protease, an enzyme long recognized as an attractive target for therapeutic intervention *vs.* AIDS. Hundreds of *in vitro* potent HIV-1 protease inhibitors have been synthesized for this indication, but most of these have suffered from a lack of bioavailability (due to rapid clearance from the circulation upon oral or intravenous delivery) that has made them otherwise unsuitable for clinical development. It is ironic that so little is known of the molecular physiology of hepatic clearance, and that no rules exist to guide the structure-based drug design of bioavailable HIV-1 protease inhibitors. In order to circumvent this problem, we have focused our drug design program on the reduction of the size of HIV-1 protease inhibitors, and have produced smaller (Mr ~500 Da) molecular weight inhibitors that retain their *in vitro* potency and antiviral activity in cells nonetheless. In addition, we find these compounds at high levels in the circulation, across a number of animal species, after oral dosing. The structure of the HIV-1 protease complex with VX-478, the lead compound in the series, will be described and compared to protease complexes with Ro 31-8959 and L-735,524, two compounds currently in clinical trials. VX-478 has completed pre-clinical development and will soon enter clinical development also.

IV. Models for protein folding and function

APPLICATION OF FUZZY SETS AND FUZZY LOGIC TO PROBLEMS IN PROTEIN SCIENCE

G.M. Maggiora,[†] J.R. Blinn,[†] K.-C. Chou,[†] D.W. Elrod,[†] T. Ross,[#] and C.-T. Zhang,[‡]
[†]Computer_Aided Drug Discovery, Upjohn Laboratories, Kalamazoo, Michigan 49007-4940, USA; [#]Department of Civil Engineering, University of New Mexico, Albuquerque, New Mexico 87131, USA; [‡]Department of Physics, Tianjin University, Tianjin, China.

Although very few applications of fuzzy sets and fuzzy logic exist in chemistry, this is not the case in engineering. There, these methods have been used to treat a broad range of problems, especially in the area of controller design where fuzzy logic has been applied to the development of controllers for a wide variety of devices from autofocus cameras, washing machines, and vacuum cleaners to complex systems for automating subways. The number of applications in business and related fields is also growing rapidly.

In contrast to traditional, or crisp, sets, where an element is either in or not in a set, the elements of fuzzy sets admit to degrees of set membership. Moreover, an element may belong to more than one set, with a different membership degree for each set. This has the effect of replacing the 'sharp boundary' of crisp sets with a smoother 'fuzzy boundary'. Such behavior is possible in the latter case since it is not necessary to specify to which set a given element uniquely belongs. 'Fuzzifying' a set is not, however, just a mathematical curiosity, for it allows one to deal with the vague and uncertain characteristics of objects or concepts. This is particularly important for molecules and molecular systems, where characteristics such as 'very soluble', 'strong base', and 'weak electron donating group' form the basis for a considerable amount of approximate chemical reasoning that is the metier of most practicing chemists.

The present talk will provide a brief introduction to fuzzy sets and fuzzy logic and will indicate their applicability to problems in protein science with two examples. The first example involves the use of fuzzy c-means clustering (FCM) to analyze the relationship of a globular protein's amino-acid composition to its folding class, *i.e.*, all-α, all-β, $\alpha+\beta$, and α/β. Fuzzy c-means clustering is a 'fuzzy variant' of normal k-means clustering, except that a given protein may appear in more than one cluster with a different membership degree in each cluster.

The second example involves the use of fuzzy logic as a means for treating conformational problems. The ability of fuzzy logic for handling highly non-linear systems such as occur in conformational analysis is supported by the many applications to non-linear dynamical systems that have already been successfully carried out (*vide supra*). The approach examined here is based upon the replacement of standard potential-energy functions by sets of fuzzy IF—THEN rules. A brief discussion of some of the ways that such fuzzy rule bases can be implemented will be presented. A potentially important

characteristic of fuzzy-logic procedures is the high degree of parallelism inherent in the evaluation of a rule set, which can lead to significant enhancements in computational speed. Another possible advantage of the fuzzy-logic approach is that it is possible to include both numerical data and linguistic information (*e.g.*, 'expert rules') in a consistent manner within the framework of the method. A brief example will be presented illustrating the development of an appropriate rule base for a model problem. Methods for 'tuning' rule sets and optimizing the performance of fuzzy-logic systems will also be briefly touched upon.

Sequence Determinants of Protein Folds

Cyros Chothia, Yehouda Harpaz and Samantha Barré
MRC Laboratory of Molecular Biology and
Cambridge Centre of Protein Engineering
Hills Road, Cambridge CB2 2QH, England

Most proteins belong to one of a relatively small number of families. The current progress in protein crystallography should soon provide, for a substantial proportion of these families, the three dimensional structure of one or more members. This means that, for the next few years, it is likely that the use of protein homologies will be the best method to determination the structures and, in many cases, the functions implicit in new sequences.

Proteins with homologous sequences have structures that are similar but not identical. Mutations in the peripheral regions of the structure can change local conformation and so produce regions where the folds are quite different. The extent of such changes is small in structures whose sequence identities are 40% or greater and variable in proteins with lower identities. In many proteins with sequence identities of 20% it is common for the peripheral regions with different folds to comprise more than half the structures. But, in some proteins with low identities, the structure of the peripheral regions are much better conserved with only 10-20% of the structures having different conformations.

This means that, if we wish to predict the structure of a protein using the known structure of a homologue with which it has a low sequence identity, it is best to find a homology in which only a few regions have different folds. We can see whether this is the case using the following procedure:

1. Identify the residues in the known structure that through packing, hydrogen bonding, or the ability to take up unusual conformations (i.e. special values of the main chain torsion angles φ, ψ, ω) are primarily responsible for the structure of the central core of the protein. Such residues are called "key" residues.
2. Carry out the same examination to identify the key residues responsible for the structure of each of the peripheral regions.
3. Align the sequence of the protein whose structure is to be predicted with the sequence of the protein of known structure.

4. Determine whether the residues at the key sites of the core regions are the same or similar in the two sequences. If they are, then do the same for each peripheral region.

Similar residues at key sites of the core region implies, of course, that the two proteins have the same basic fold. The extent of the structural similarities will depend upon the extent to which the key residues in the peripheral regions also match.

We have used this procedure to investigate:

1. the evolutionary history of the structure of immunoglobulin binding sites;
2. the structure of the neural cell adhesion molecules (NCAMs), the cell surface proteins that play a central role in the regulation, organisation and maintenance of interactions made by embryonic and adult neurones and their target tissues, and
3. the structural role of the natural mutations that occur in the cell adhesion molecule L1 and which are responsible for X-linked hydrocephalus.

The results of this work will be described.

Limited Folding in Proteins

(*I.Fita*.. CSIC-UPC. Diagonal 647. 08028 Barcelona).

The problem of predicting the structure of a protein from its amino acid sequence constitutes the classical "protein folding problem". A similar but simpler problem is that of predicting the stability and folding of a protein when a closely related structure is known. This situation is often found when studying the formation of macromolecular complexes or when a mutation, or a chemical modification, is introduced in a protein whose three-dimensional structure was known. It also appears, as a major practical problem, during the refinement process of protein structures empirically determined.

X-ray crystallography (together with NMR spectroscopy) is the most powerfull tool available to provide detailed three-dimensional structural information of macromolecules. The last years have seen a technical revolution in a number of critical stages in solving a protein structure. These include, for example, methods to obtain accurate initial models (1) that are then routinely improved by crystallographic refinement. A number of algorithms have been described for this purpose. These include the use in the refinement of explicit molecular-mechanics force fields (2), model constraints and restraints (3,4) and, more recently, force-field-based molecular-dynamics (5,6).

In a given protein structure solvent-atom and atom-atom nonbonded interactions can be quantitatively described using accessible (7) and contact (8) atomic surface areas, respectively. This description has been incorporated as new restraints in the refinement program of Hendrickson&Konnert (3). The algorithm allows then to calculate and, if required, to optimize the extent of atomic areas in a molecule that are accesible to the solvent or in contact with non-solvent atoms. Information on the correctness of atomic interactions in a given configuration can thus be used either as added data during the crystallographic refinement or,

independently, in the analysis and modelling of molecular structures.

The physical implications of this unified description in which solvent-molecule and the molecular non-covalent interactions are formally treated equal will be discussed.

References

(1) T.A.Jones *et al.*, (1991), Acta Cryst., **A47**, 110-119.

(2) A.Jack&M.Levitt, (1978), Acta Cryst., **A34**, 931-935.

(3) W.A.Hendrickson&J.Konnert, (1985), Methods Enzymol. **115**, 252-270.

(4) J.L.Sussmann et al., (1977), Acta Cryst., **A23**, 800-804.

(5) A.T.Brünguer,J.Kuriyan&M.Karplus, (1987), Science, **235**, 458-460.

(6) M.Fujinaga,P.Gros&W.F.VanGunsteren,(1989), J.Appl.Cryst., **22**, 1-8.

(7) B.Lee&F.M.Richards, (1971), J.Mol.Biol., **55**, 379-400.

(8) C.Chothia, (1974), Nature, **254**, 304-305.

Protein's conformational space studies using perturbation-relaxation MD methods

O.Tapia

A perturbation-relaxation method is proposed to sense the response of MD-equilibrated or model built structures to external perturbing conditions. The procedure permits searching conformations that may be interconvertible with a given structure when the external perturbation is switched off. Finding conformations that are interconvertible with a given structure when external conditions are changed is an interesting problem that can be treated by molecular dynamics simulations. The changes in the solution properties may affect the cohesive forces of proteins which, in the limit, may lead to denaturing of the respective fold.

The application of a perturbation relaxation procedure to a simulated system in full aqueous representation has a practical drawback in that the relaxation times are often too long to yield a relatively rapid tool for exploring the conformational space. One way to go around the slow relaxation time problem is to use a collisionless solvent model (in previous work we have named it as non-inertial solvent (NIS) model). This is GROMOS' parameter set 37D4. This scheme has the advantage of avoiding explicit inclusion of water and couterions. The latter are assumed to counterbalance the charges of ionizable residues; thus, the protein is globally electroneutral. As collisions between protein atoms and the solvent are not present, the time scale for relaxation from non-equilibrium situations is bound to be shorter than for full solvent models. GROMOS–NIS constitutes a sort of equilibrium simulation engine. The perturbation is entered into the program via a calculation of the solvent accessible surface during the perturbing run. The program GEPOL has been adapted for this purpose. To each atom having solvent accessibility, we register the normal ($\mathbf{n}(k)$) to the accessible surface and its value (S_k). The perturbation is built as follows: for each surface atom the sum of the intramolecular van der Waals and electrostatic force is stored ($\mathbf{A}_{intra}(k)$) we assume that the solution acts on the protein via a force acting at its surface atoms

$$\sigma(k) = -Bk\ \mathbf{A}_{intra}(k)\cdot\mathbf{n}(k)\ \ \mathbf{n}(k)$$

where Bk in one class of simulations was given a unit value for all atoms, while in a second class of simulations it was given the value Sk.

The response of this system to an external perturbation, that would mimick changes introduced by solution properties (denaturing conditions), is determined by using molecular dynamics simulations with time dependent perturbation forces

$$\mathbf{A}(k,t) = (1-\lambda(t))\ \mathbf{A}_{intra}(k)_{NIS} + \lambda(t)\ \sigma(k)$$

where λ is an auxiliary time–dependent variable that goes from 0 to 1. This is similar to the technique used in free energy calculations, although thermodynamic equilibrium is not

assumed here. This particular procedure will be referred to as "solution-like simulation".

Zinc finger protein Zif-268, which is used as an example here, is one of the major structural motifs involved in eukaryotic protein-nucleic acid interactions. Pavletich and Pabo reported the crystal structure of a Zif268-DNA complex at 2.1Å resolution and refined to a crystallographic R-factor of 18%. This structure provides a framework for understanding how zinc fingers recognize DNA and, as suggested by these authors, it may provide a basis for the design of novel DNA-binding proteins.

The following set of simulations will be discussed: i) Zinc-free Zif-268 in pseudo vacuum and solution-like conditions ranging from mild up to denaturing conditions; ii) Zif-268 in pseudo vacuum and solvent-like conditions; iii) Full water Zif-268 and full water plus solution-like conditions.

The simulated system shows a behavior under varying perturbing conditions which can be correlated with the ability shown by the real system to bind DNA. The role of zinc is determining the tertiary structural arrangement, but seems to have a minor effect on the presence of secondary structural elements. Zinc appears then as the determining factor in selecting the conformational space apt to bind DNA. The structure connected to the conformational space found in the X-ray complex appears, in the simulations, as a subspace related to the unfolding of the system, while the structure in pseudo-vacuum has a tertiary arrangement inapt to bind DNA with the same structure as in the real crystal. If anything, such behavior may be related to a highly structured "molten globule". This is at least the picture drawn from the pseudo-vacuum simulations. More experimental data is necesary to interpret the actual folded state of the protein alone. If the NIS scheme actually reflects the crystal structure of the protein alone, this work predicts that such a crystal structure would differ in tertiary structure when compared to the DNA-Zif-268 complex. If such a conclusion were to be confirmed by experiment (with an X-ray structure of the protein alone), the present results would suggest that this type of proteins have to be set in a quasi-denaturing situation in order to achieve the functionally correct tertiary structure as it is found in Zif-DNA crystal.

Simple lattice models of protein folding

Ken A. Dill, UCSF, San Francisco, CA, 94143

The model

We have studied a "simplest" model of proteins as self-avoiding chains configured on lattices. We assume that the dominant force of folding is due to hydrophobic interactions. Hydrophobic interactions are nonlocal in the sequence. We suppose each protein is a specific sequence of hydrophobic (H) and polar or ionic (P) monomers. When two H monomers are adjacent in space, it is weighted by a favorable Boltzmann factor. By exhaustive enumeration and other methods, we have explored the lowest energy configurations of such chains, for all possible HP sequences, and we have studied the kinetics of folding by Monte Carlo methods. We have developed fast search methods to find lowest energy states for 3-dimensional chains up to 88-mers. These methods are reviewed in (1-5).

Protein Structures

This simple HP model shows structures characteristic of proteins. When the HH attraction is weak, molecules adopt an ensemble of denatured configurations. With increasing HH attraction, the chains undergo a cooperative folding transition, involving a peak in the specific heat, to compact structures that have H monomers buried into a hydrophobic core. We find that the collapse process to compact states, stabilizes secondary structures. There are very few "ground states". Certain sequences tend to have tertiary structures - helical bundles, $\alpha\beta$ barrels, the parallel β helix - resembling those of proteins. The special sequences with these properties are those designed to satisfy two goals: (i) positive design - to fold to the desired native conformation, and (ii) negative design - to avoid folding to all other conformations.

Cooperativity

We use this model and a related simple exact model of sidechains to explore the basis for 2-state cooperativity in protein folding. It is known that helix- coil cooperativity, which follows the behavior of the 1 dimensional Ising model, is not 2-state. We find that certain HP sequences do have 2-state cooperativity of folding because of the development of a hydrophobic core. Our simplest model of sidechains shows no evidence for a sidechain unlocking basis for 2-state behavior. Nor do homopolymer collapse experiments, in which sidechains freeze upon collapse, show evidence for 2-state cooperativity.

Folding Kinetics

The HP model has two kinetic stages of folding: (i) a fast collapse process of hydrophobic "zipping", and (ii) a slow process of barrier-climbing, breaking the incorrect contacts formed in the fast collapse. The transition state for folding involves an an opening up of the chain, which allows it to repair mistaken contacts.

References

(1) H. S. Chan and K. A. Dill, Physics Today, 46, 24, 1993

(2) H. S. Chan and K. A. Dill, Ann Rev Biophys Biophys Chem., 20, 447, 1991.

(3) K. A. Dill, Biochemistry, 29, 7133, 1990.

(4) K. Yue and K. A. Dill, Phys. Rev. E, 48, 2267, 1993.

(5) K. A. Dill, K. M. Fiebig, H. S. Chan, Proc. Natl. Acad. Sci., 90, 1942, 1993.

Computer Simulation of Protein Folding and Unfolding

Michael Levitt

Department of Structural Biology
Stanford Medical School
Stanford, CA 94305

October 1994

We present results of computer simulations of (a) protein folding using simplified models at different resolutions and (b) protein unfolding using detailed all-atom models that include solvent and ions.

Simplified models have played an important role in attempts to model protein folding by computer simulation: By eliminating much of the detail, such models are computationally more tractable. While such models are useful, a central question is whether there is sufficient detail in a particular model to make the native fold stand out from the vast number of alternate non-native folds. More specifically, one hopes that the native fold has a very low free-energy value.

Here we use a simplified representation of proteins that allows all possible conformations to be generated. We then search this conformational space in order to find those folds that have low energies with a simple pair-wise residues contact potential. Results on a wide variety of small proteins show that, even at this low resolution, there is significant selectivity. Using the correct sequence of the protein, selects for those folds that are more similar to the native conformation of the protein. Using a shuffled sequence does not show such selectivity.

The second part of the talk will describe simulation of the first steps of thermally induced unfolding of a small protein (bovine pancreatic trypsin inhibitor) using a detailed all-atom model with solvent and counter-ions. This work, is based on accurate molecular dynamics simulations that conserve energy and use periodic boundary conditions. Simulations run for 500 picoseconds at room temperature remain native-like. Simulations at 100 C or 150 C lead to unfolding to a compact globular state with properties close to those observed experimentally for the molten globule state. These properties, include a residual hydrophobic core, a large amount of secondary structure, high side-chain mobility, and a "dry" interior. The thermal unfolding of simple peptides and an entire protein in water is illustrated dramatically with a 16mm movie.

Protein Folding Simulations: From the Native to the Denatured State and Back Again

Martin Karplus

Department of Chemistry
Harvard University
Cambridge, Massachusetts 02138
U.S.A.

and

Laboratoire de Chimie Biophysique
Institut le Bel
Universite Louis Pasteur
67000 Strasbourg FRANCE

The mechanism of protein folding is one of the areas of structural biology where simulations can play an important role. For different aspects of the problem (e.g. unfolding from the native state to the compact globule, folding from the denatured state to an organized globule) different techniques are most useful.

The transition in barnase from the native state to a compact globule has been studied with high temperature all-atom molecular dynamics simulations in the presence of solvent. A partial destruction of the α-helices and the outer strands of the β-sheet is observed with water molecules replacing the hydrogen bonds of the secondary structural elements. Simultaneously, the main α-helix moves away from the β-sheet and exposes the principal hydrophobic core, many of whose nonpolar sidechains, beginning with the ones near the surface, become solvated by hydrogen-bonded water molecules. This step involves a significant increase in the solvent-exposed surface area; the resulting loss of stability due to the hydrophibic effect may be the major source of the activation barrier in the unfolding reaction. The detailed mechanism for the first stage of the denaturation of barnase,

including the essential role of water molecules, is likely to be representative of protein denaturation, in general.[1,2]

The number of all possible conformations of a polypeptide chain is too large to be sampled exhaustively. Nevertheless, protein sequences do fold into unique native states in seconds (the Levinthal paradox). to determine how the Levinthal paradox is resolved, we use a simplified lattice Monte Carlo model in which the global minimum (native state) is known. The necessary and sufficient condition for folding in this model is that the nativ estate be a pronounced global minimum on the potential surface. This guarantees thermodynamic stability of the native state at a temperature where the chain does not get trapped in local minima. Folding starts by a rapid collapse from a random-coil state to a random-semi-compact globule. It then proceeds by a slow,rate-determining search through the semi-compact states to find a transition state from which the chain folds rapidly to the native state. The elements of the folding mechanism that lead to the resolution of the Levinthal paradox are the reduced number of conformations that need to be searched in the semi-compact globule ($\sim 10^{10}$ versus $\sim 10^{16}$ for the random coil) and the existence of many ($\sim 10^{3}$) transition states. The results have evolutionary implications and suggest principles for the folding of real proteins.[3,4]

References:

1. Molecular Dynamics Simulation of Protein Denaturation: Solvation of the Hydrophobic Cores and Secondary Structure of Barnase, *Proc. Natl. Acad. Sci. USA* **91**, 1746-1750 (1994), by A. Caflisch and M. Karplus.

2. Molecular Dynamics Studies of Protein and Peptide Folding and Unfolding, in *The Protein Folding Problem and Tertiary Structure Prediciton*, K. Merz, Jr. and S. Le Grand, Editors, Chapter 7, 195-232 (Birkhäuser, Boston, MA, 1994), by A. Caflisch and M. Karplus.

3. Kinetics of Protein Folding: A Lattice Model Study of the Requirements for Folding to the Native State, *J. Mol. Biol.* **235**, 1614-1636 (1994), by A. Šali, E. Shakhnovich, and M. Karplus.

4. How Does a Protein Fold? Nature **369**, 248-251 (1994), by A. Šali, E. Shakhnovich, and M. Karplus.

POSTERS

Sequence divergence analysis as a tool for GPCR modeling

Ibon Alkorta[1] and Ping Du[2]

[1] Instituto de Quimica Medica, Juan de la Cierva-3. 28006-Madrid. Spain.
[2] Synaptic Pharmaceutical Corporation. 215 College Road. Paramus, NJ 07652. USA

Sequence divergence analysis is a newly developed method for three dimensional structure prediction of proteins. This method is based on the observed mutabilities tendencies of the aminoacids as a function of its environment. Based in this idea, a variability parameter has been developed that allows to identify the solvent and lipid accessible residues using the sequences of a family of homologous sequences. The variability has been used to predict the transmembrane regions, secondary structure, and bundle shape of the receptor, allowing us to build a 3-D model of the family in question.

The method was validated with the bacteriorodopsin. In addition, the structure of the human rhodopsin and dopamine D1 receptors were constructed using the same methodology. In the case of the human rhodopsin, 42 sequences of opsins and visual pigments were used. The model obtained is consistent with experimental mutation data and the low resolution electron density map of bovine rhodopsin recently published. For the dopamine D1 receptor, 92 sequences of amine neurotransmitter receptors have been used. A ligand binding site has been proposed that agree with a recent pharmacophore propose for this receptor.

DESIGN AND BIOPRODUCTION OF OPTIMIZED CHELATING PEPTIDES TO REMOVE HEAVY METALS FROM WATER.

Atrian, S.*; Oliva, B.#; Cols, N.*; Gonzàlez-Duarte, R.*; Duran, N.§; Romero, N.§ and Gonzàlez-Duarte, P.§.

*Departament de Genètica. Facultat de Biologia. Universitat de Barcelona. Av. Diagonal 645. 08071-Barcelona, Spain.
#Institut de Biologia Fonamental. Universitat Autònoma de Barcelona. 08193-Bellaterra, Barcelona, Spain.
§Departament de Química. Facultat de Ciències. Universitat Autònoma de Barcelona. 08193-Bellaterra. Barcelona, Spain.

The aim of our project is the design and bioproduction of polypeptides with high metal chelating ability to be used in water treatment and detoxification. Metallothioneins (MT) are cysteine-rich, low molecular-weight proteins which protect against metal toxicity in the animal kingdom. This feature makes them an attractive model for the design of chelating polypeptides, but bioproduction and handling of the native proteins is limited by their high cysteine content. New polypeptides with structural similarities to highly conserved metallothionein fragments were designed to avoid oxidation problems and maintain their coordination abilities.

Molecular modelling of the ß-fragment of mouse metallothionein revealed which residue -aspartic acid, glutamic acid or histidine- introduced the fewest structural changes when substituting terminal cysteine residues (Cys-5, Cys-13, Cys-19, Cys-21). New mutant forms were modelled from crystal data of the Cd_5, Zn_2-mouse MT-II. These studies were carried out in a Crimson-Elan work station of Silicon Graphics with the program TURBO-FRODO, which allows the replacement of one or more residues of a polypeptide chain without altering its native main chain conformation. This is achieved by minimizing the van der Waals interactions of the new lateral chains, whose orientation is calculated from crystallographic data. Nevertheless, the program does not determine whether the orientation of the new residue is suitable for the maintenance of coordination to the metallic centre. Thus, the torsion angles of the lateral chain have to be altered in order to establish this coordination. In the limiting case in which this torsion involves breaking the bonds of the main chain, the substitution is ruled out. Our results can be summarized as follows: when only one Cys residue was substituted, structural changes increased accordingly to the sequence Asp < His < Glu for Cys-13, Cys-19 and Cys-21, while for Cys-5 the order was Asp < Glu < His; if more than one Cys residue was substituted, the global effect was the sum of the individual effects, except for Cys-5 and Cys-21, both of which coordinate to the same metal atom. Here it is observed that: (1) Cys-21 can be substituted only by Asp if Cys-5 is replaced by His, (2) the most favourable combination is [Asp 5, Asp 21] and (3) the most unfavourable substitution is [His 5, His 21].

On the other hand, two chemically-synthesized 15-residue peptides, closely related to mouse and *Agaricus bisporus* MT fragments, have been spectrophotometrically titrated with Co^{2+}. The number of bands together with the position of their maximum absorption provided data on the coordination geometry and the nature of donor atoms.

The synthesis of the optimized polypeptide chains will be approached through cloning and expression of the DNA coding sequence. Thus, three heterologous systems were set up in *E.coli* to express the mouse MT-I gene as a chimeric product. Constitutive (P_{SPA}) or inducible (P_{tac}, λPR) promoters control the synthesis of three MT fusion proteins. In one case, a signal peptide directs the product to the extra cellular medium. The recombinant proteins were purified, characterized and assayed by atomic absorption (ICPAES) for heavy metal affinity.

Optimal protein structure alignments by multiple linkage clustering : Application to the ubiquitin family, to the Greek key group and to the analysis of conformational changes in citrate synthase.

Nathalie S. Boutonnet[1,2], Marianne J. Rooman[1] & Shoshana J. Wodak[1]

[1] Unité de Conformation des Macromolécules Biologiques, Université Libre de Bruxelles, CP 160/16, Av. P. Héger, 1050 Bruxelles, Belgium.

[2] Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, Paris 75005, France.

A fully automatic procedure for aligning two protein structures is presented. It uses no prior information on residue equivalences, relies on the root mean square deviation of all backbone atoms (N, Cα, C and O) after superposition as the sole similarity measure and is designed to yield optimal solutions. In a first step, the procedure identifies protein segments with similar conformations in both proteins. These segments correspond either to secondary structure elements, or to 5-residues fragments. In a second step, a novel multiple linkage clustering algorithm is used to identify segment combinations which yield optimal global structure alignments. This algorithm is a generalized version of the tree-like clustering method. Its original feature is that the same segments can be joined more than once, thereby defining multiple intertwined clustering trees, from which several distinct structure alignments can usually be obtained for a given pair of protein. These multiple solutions are exploited to automatically define structural motifs common to a protein family. Furthermore, an automatic analysis of the clustering trees is developped which enables to identify rigid-body movements such as those occurring between domains or secondary structures. To illustrate the performance of our procedure, we apply it first to determine the consensus structure motif of families of distantly related proteins. One groups the three α+β proteins ubiquitin, domain B1 of protein G and ferredoxin. The other family represents β-proteins from the Greek key group, in particular actinoxanthins, immunoglobulin variable domains and plastocyanins. In a second application, our algorithm is applied to identify the rigid-body movements that occur between the unliganded and liganded forms of citrate synthase.

Georg Casari,
EMBL, Meyerhofstrasse 1
HEIDELBERG D-69117 (Germany)

FUNCTIONAL RESIDUES IN PROTEIN SEQUENCE SPACE

The biological activity of a protein typically depends on the presence of a small number of functional residues placed at precise positions in a structural framework. In the absence of detailed experimental information it would therefore be very useful to be able to identify such key residues from the amino acid sequence alone. Classically, strictly conserved residues are predicted to be involved in function, but often conservation patterns are more complicated, e.g., when a subset of proteins in family adapts to specific functional constraints. Here, we present a novel method that exploits such subtle patterns of conservation for the prediction of functional residues. The key invention of the method is a simple but powerful representation in which entire proteins as well as combinations of residues at particular positions are vectors in a generalised 'sequence space'. Projection of these vectors onto a lower dimensional space using principal component analysis reveals groups of residues specific for particular subfamilies. It is such specificity-determining residues that are predicted to be directly involved in protein function. Proof of the concept comes from control examples in which sets of functional residues are known experimentally, e.g., the switch II region of Ras p21 oncogene protein and the specific binding pocket of the phosphotyrosine-binding SH2 domains. Based on the new method we present testable predictions for sets of functional residues in SH2 domains.

STRUCTURAL AND FUNCTIONAL ANALYSIS OF PROTEIN TRWC, INVOLVED IN PLASMID R388 CONJUGATION

Matxalen Llosa, Guadalupe Grandoso, Juan Carlos Zabala and Fernando de la Cruz

Departamento de Biología Molecular, Universidad de Cantabria, C. Herrera Oria s/n, 39011 Santander.

TrwC is an essential protein in plasmid R388 conjugation. It shows two independent enzymatic activities: DNA helicase, and DNA-strand transferase. Its function in conjugation is most probably tripple: it nicks a specific site in a specific DNA strand on the origin of conjugal transfer (the nic site), completely unwinds the nicked strand, and religates the nicked DNA at the end of the unwinding reaction.

The DNA-strand transferase activity is localized to the 300 N-terminal amino acids. The sequence of this domain shows three conserved motifs, which are shared by a family of proteins involved in similar nicking reactions in other plasmids. Two of these motifs are also shared by a much broder range of DNA-strand transferases, implicated in rolling circle replication of plasmids and phages, a mechanism with obvious similarities with conjugative DNA replication. The reaction mechanism involves the formation of a binary complex by covalent linkage between a protein tyrosyl residue, and the 5' end of the nic site. The 3' end of the nick remains also tightly bound to the complex, since the DNA superhelicity is maintained, and the 3' end is not accessible to DNA polymerases. After the unwinding reaction (carried out by the helicase domain), TrwC can reverse the nicking reaction, recircularizing the displaced DNA strand, and giving rise to a circular ssDNA intermediate. The enzyme can also transfer the DNA strand bound to it, to a free 3'-OH containing any DNA sequence exactly matching the 5' end of the nic site.

The DNA helicase activity resides in the C-terminal half of the protein. In this case there is a series of seven conserved motifs, diagnostic of a family of DNA helicases, exemplified by proteins Rep and UvrD of E.coli. The mechanism of reaction of helicases is a very active and as yet unresolved subject of scientific analysis.

The resolution of the three dimensional structure of TrwC would provide a more adequate comprehension of these two mechanisms of reaction, essential in a wide range of biochemical processes.

ON HOW SMALL CHANGES IN FORCE FIELD PARAMETERS AFFECT THE SIMULATED DYNAMICS OF SOLVATED PROTEINS. ANALYSING THE PHYSICAL PROPERTIES OF THE POTATO CARBOXYPEPTIDASE INHIBITOR (PCI) IN SOLUTION.

X. Daura, B. Oliva, E. Querol, F.X. Avilés and O.Tapia†

Institut de Biologia Fonamental; Universitat Autònoma de Barcelona; E-08193 Bellaterra, España.
†Fysikalisk Kemiska Institutionen; Uppsala Universitet; S-751 21 Uppsala, Sverige.

Recently, partial modifications to the GROMOS87 force field parameter set [1] have been proposed by different authors in order to improve the convergence between the expected and the observed physical properties of solvated systems and interfaces. These modifications have been mainly made on the van der Waals C_{12} repulsion parameters accounting for interactions between any oxygen and carbon types, leading to a higher repulsion between aliphatic or aromatic groups and surrounding water molecules. In this report we test five slightly different parameter sets, which we call psG [1], psA [2], psW [van Gunsteren, personal communication], psD [Daura, unpublished results], and psB [3], by analysing the water solvation effects on a small protein. The purposes of this molecular dynamics (MD) study are to evaluate the sensitivity of systems describing protein-water interactions to the van der Waals and electrostatic parameters, and to determine which of the sets better describes the expected physical properties of these systems. PCI is particularly suitable for MD studies because it is small (39 residues), and its structure is known both in aqueous solution [4] and in crystal complex with carboxypeptidase A [5].

For the present study we have used the GROMOS package. Five MD simulations, called SimG, SimA, SimW, SimD, and SimB, were seed with the parameter sets above mentioned and the SPC bulk water model [6]. A sixth MD simulation, SimB/E, was seed with the parameter set psB and the SPC/E bulk water [7], slightly more polar than the original SPC. The crystallographic structure of PCI was chosen in all six cases to define the initial system, completed with 1259 water molecules and 8 counter ions. The non-polar hydrogen atoms were treated by the united atom approach. After a total of 1800 steps of steepest descent energy minimisation, the six 500 picoseconds (ps) MD simulations were carried out at constant temperature (293 K) and pressure (1 atm) under periodic boundary conditions, using an integration time step of 0.002 ps and constraining bond lengths to equilibrium values using the SHAKE algorithm. A twin-range cut-off for non-bonded interactions of 0.8/1.3 nm was used, being the pair list and forces for the range between 0.8 and 1.3 nm updated every 10 time steps.

An extensive analysis of the yielding trajectories shows that the simulation of the dynamics of solvated proteins is very sensitive to the van der Waals parameters chosen for the C-O interactions as well as to the dipole moment of the water molecule. From comparison to the crystallographic and NMR data and to the expected behaviour of a number of physical properties, the best dynamics are obtained in simulations SimW and SimD. The rest of the simulations show larger deviations from both the reference data and the expected physical behaviour.

REFERENCES

1. van Gunsteren W.F., Berendsen H.J.C., Biomos b.v. Groningen, 1987.
2. Åqvist J., Medina C., Samuelson J.-E., Protein Engineering, 1994, 7: 385-391.
3. van Buuren A.R., Marrink S.-J., Berendsen H.J.C., J. Phys. Chem., 1993, 97: 9206-9212.
4. Clore G.M., Gronenborn A.M., Nilges M., Ryan C.A., Biochemistry, 1987, 26: 8012-8023.
5. Rees D.C., Lipscomb W.N., J. Mol. Biol., 1982, 160: 475-498.
6. Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., Hermans J., Intermolecular Forces.; Pullman B., Ed.; Reidel: Dordrecht, 1981; pp. 331-342.
7. Berendsen H.J.C., Grigera J.R., Straatsma J.P., J. Phys. Chem., 1987, 91: 6269.

Analysis of Sequence and Structure Relationships in the Calycin Protein Superfamily

DARREN R FLOWER

Department of Physical Chemistry
Fisons Plc, Pharmaceuticals Division
Research & Development Labs, Bakewell Road
Loughborough, Leicestershire, UK, LE11 ORH

The Calycin protein superfamily, an example of a so-called "Structural Superfamily" where obvious similarities of three-dimensional structure are not easily discernible at the sequence level, is composed of three distinct families of ligand binding proteins: the Lipocalins, a highly diverse family of extracellular transport proteins; the Fatty Acid-Binding Proteins (FABPs), a family of predominantly intracellular proteins involved in lipid metabolism; and the Avidins, proteins with a remarkable affinity for the vitamin biotin which have found important applications in biotechnology. The Calycin superfamily is characterized by a similar overall folding pattern (an antiparallel β-barrel, with a repeated +1 topology, possessed of an internal ligand binding site), within which large parts of the lipocalin, FABP, and avidin structures can be structurally equivalenced. The first structurally conserved region within the three-dimensional alignment, or common core, characteristic of the three families corresponds to an unusual structural feature (a short 3_{10} helix leading into a β-strand, the first of the barrel), conserved in both its conformation and its location within their folds, which also displays characteristic sequence conservation.

In this poster relationships of sequence and structure, within this fascinating group of proteins, are explored. New insights into the sequence determinants of the Calycin fold are presented as is a detailed analysis of sequence relationships within the various families and subfamilies which go to make up the superfamily.

Recognition of distantly related proteins through energy calculations

Ruben Abagyan, Dmitrij Frishman, and Patrick Argos

European Molecular Biology Laboratory, Postfach 102209, Meyerhofstr., 1, 69012 Heidelberg, Germany

We have developed a new method to detect remote relationships between protein sequences and known three dimensional structures based on direct energy calculations and without reliance on databank statistics. The likelihood of a residue to occupy a given position on the structural template was represented by an estimate of the stabilization free energy made after explicit prediction of the substituted sidechain conformation. The underlying assumption of our approach is that proteins are permissive to minor structural rearrangements and that the behaviour of the protein interior can be approximated by that of a non-polar liquid. The profile matrix derived from these energy values and modified by increasing the residue self-exhange values successfully predicted compatibility of heat-shock protein and globin sequences with the three-dimensional structures of actin and phycocyanin, respectively, from a full protein sequence databank search. The high sensitivity of the method makes it a unique tool for predicting the three-dimensional fold for the rapidly growing number of protein sequences.

Circular Dichroism Calculations using Molecular Dynamics Trajectories of Isolated Helices of Myoglobin

Jonathan D. Hirst and Charles L. Brooks III
Department of Chemistry, Carnegie Mellon University,
Pittsburgh, PA 15213

The apo form of myoglobin has two non-native stable states that have been experimentally characterized. Investigation of these states has suggested possible folding pathways for myoglobin. Previously, molecular dynamics simulations of the "native" form of apomyoglobin have provided further data on this system. We report molecular dynamics simulations on solvated isolated helices of myoglobin. Analysis of these simulations allows us to study the relative stabilities of the helices of myoglobin and to examine the importance of tertiary interactions. We have predicted the time-evolution of the circular dichroism of the isolated helices of myoglobin, using empirical and semi-empirical methods. We examine the potential of these calculations as an tool for interpreting time-resolved circular dichroism experiments of folding transitions of helices and proteins.

UNDERSTANDING OF THE SOLVENT POLARIZATION: IMPLICATIONS IN MODELLING STUDIES

F. J. Luque[a] and M. Orozco[b]

a Departament de Farmàcia, Unitat Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, 08028 Barcelona, SPAIN

b Departament de Bioquímica i Fisiologia, Facultat de Química, Universitat de Barcelona, 08028 Barcelona, SPAIN

The solvent plays a critical role in the determination of the thermodynamics and kinetics of chemical processes, among which the conformational changes and intermolecular interactions in macromolecular systems are of particular interest. The simulation of solvent effects can be achieved with reasonable accuracy by means of force-field methods, self-consistent reaction field techniques and mixed Quantum Mechanics-Molecular Mechanics strategies. Force-field derived methods are widely used to study the structure and dynamics of proteins and nucleic acids in solution. However, despite the popularity of these methods, the mutual solute-solvent polarization is not explicitly included. This effect, which is especially relevant for polar solvents, induces a charge redistribution of the electron density, which, in turn, influences the molecular reactivity.

A reliable simulation of biomolecular interactions would require the use of solute Hamiltonians enough flexible as to incorporate the solvent influence. In this context SCRF methods can be extremely helpful in the understanding of the solvent polarization. Different aspects ranging from the thermodynamics of the solvation process, the change in the molecular properties arising from the charge redistribution induced upon solvation and the implications in the force-field parametrization are presented. Finally, a procedure to include solvent effects in an implicit way using SCRF results is discussed.

A Comparison of Particle-Particle Particle-Mesh and Ewald Methods for Calculating Electrostatic Interactions in Periodic Molecular Systems.

Brock A. Luty[1], Malcolm E. Davis[2], Ilario G. Tironi[1] and Wilfred F. van Gunsteren[1]

[1] Laboratorium für Physikalische Chemie
ETH Zentrum
CH-8092 Zürich
Switzerland

[2] Bristol-Myers Squibb Pharmaceutical Research Institute
P.O.Box 4000
Princeton, NJ 08543-4000
USA

Abstract

Because of the long-range nature of Coulombic interactions, electrostatics plays an important role in chemical and biochemical processes. Evaluation of the electrostatic interaction in molecular simulations is a difficult and computationally demanding task. If, as is commonly the situation, the molecular systems is simulated using periodic boundary conditions, then the electrostatic interactions can be evaluated using lattice-sum methods. We compare the Particle-Particle Particle-Mesh (PPPM) and Ewald methods for calculating electrostatic interactions in periodic molecular systems. A brief comparison of the theories shows that the methods are very similar differing mainly in the technique which is used to perform the "k-space" or mesh calculation. Because the PPPM utilizes the highly efficient numerical Fast Fourier Transform (FFT) method it requires significantly less computational effort than the Ewald method and scales almost linearly with system size.

Genetic algorithm-based protein structure comparisons: pairwise and multiple superpositions.

Alex C.W. May and Mark S.Johnson
Imperial Cancer Research Fund Unit of Structural Molecular Biology,
Department of Crystallography,
Birkbeck College,
University of London,
Malet Street,
London WC1E 7HX,
United Kingdom.

We have developed a genetic algorithm-based procedure for rigid structure superposition, in which the assignment of initial equivalences is not required [1]. A genetic algorithm (GA) is used to search for a near optimal solution of the rigid-body superposition of two whole protein structures. The GA maintains a population of superpositions. Bit strings are used to represent the transformations (translation along and rotation about each of the three axes) applied to one of the structures to obtain a pairwise superposition. The GA applies a fitness function to assess the success of the encoded transformations. Once the quality of the trial superpositions has been evaluated, the best superpositions survive to mate in the GA and make better superpositions which replace poorer ones. The selection of topological equivalences in the final structure alignment is by dynamic programming. Dynamic programming is also used to evaluate trial superpositions represented in the bit strings. Finally, a least-squares fitting algorithm is used to optimize the fit between the GA-matched equivalences.

We have incorporated an improved translation operation and a method to select one sequence alignment to represent a superposition into the GA-based structure alignment method [2]. In addition, we describe extension of the method to comparison of two or more 3-D structures by adapting an approach for the multiple alignment of protein sequences. The method has been tested on members of different protein families.

1. May ACW, Johnson MS. 1994. Protein structure comparisons using a combination of a genetic algorithm, dynamic programming and least-squares minimization. Protein Eng. 7: 475-485.

2. May ACW, Johnson MS. 1994. Submitted.

THREE DIMENSIONAL STRUCTURE OF A L.M.W.-NAPIN

RAFAEL MONSALVE CLEMENTE
Dpt. de Bioquímica y Biología Molecular
Facultad de Química. Universidad Complutense
Ciudad Universitaria. 28040 MADRID (Spain)

The 3D structure of a low molecular napin is about to be known. Napins are a huge family of seed storage proteins of dicotyledoneous plants, which have been extensively studied in the last ten years. Apart from their commercial and nutritional importance, some members of this family are known to be major allergens in certain species (yellow and oriental mustard), and other members have also been described as proteinase inhibitors, and, very recently, as natural antifungal proteins. Therefore, its potential functions are going to be studied more precisely when its 3D structure will be known.

The structural and immunological features determined in our laboratory, as well as computer-predicted structural data (secondary structure, antigenicity, flexibility, etc.), will be discussed and compared with the preliminary 3D data obtained by NMR spectroscopy.

How does a β hairpin denaturate?
study of the thermal unfolding of an excised peptide from barnase.

Luisa Pugliese, Martine Prévost and Shoshana Wodak.
Unité de Conformation des Macromolécules Biologiques. CP160/16
Université Libre de Bruxelles, Av. P. Héger, 1050 Brussels, Belgium.

Several molecular dynamics simulations have already been performed to study protein denaturation (1, 2, 3, 4). However no detailed analysis has, to our knowledge, been reported on the unfolding of a β–hairpin in solution, using molecular dynamics simulations.

The model system is the peptide 85–102 from barnase, a small extracellular ribonuclease from *Bacillus amyloliquefaciens* (5). It was shown (6) that this peptide adopts its native structure in the folding intermediate of barnase, when all the tertiary interactions are not yet present. This indicates that it would be a potentially stable conformation when excised from the protein.

To simulate unfolding, a series of 300 ps molecular dynamics simulations were carried out at increasing temperatures: 300 K, 450 K and 600 K. From the generated trajectories the evolution of the atomic interactions during the denaturation process was analyzed.

Simulations of the whole protein were also performed at 300 K and 450 K.

Comparison between the behaviour of the β hairpin in solution and inside the protein from which it was excised, will be also reported in the poster.

References:

(1) Mark, A.,E. and W. Van Gunsteren
Biochemistry (1992) 31, 7745–7748.
(2) Daggett, V. and Levitt, M.
Proc. Natl. Acad. Sci. USA (1992) 89, 5142–5146.
(3) Daggett, V. and Levitt, M.
J. Mol. Biol. (1993) 232, 600–619.
(4) Caflisch, A. and Karplus, M.
Proc. Natl. Acad. Sci. USA (1994) 91, 1746–1750.
(5) Maugen, Y., Hartley,R.W., Dodson,E.J., Dodson, G.G., Bricogne, G., Clothia, C. and Jack, A.
Nature (1982) 297, 162–164.
(6) Matouschek,A., Serrano, L. & Fersht, A.R.
J. Mol. Biol. (1992) 224, 819–835.

COMPUTER SIMULATION OF PROTEIN UNFOLDING *IN VACUO*

Angel R. Ortiz[†#], Rebecca C. Wade[†]

[†] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany.

[#] Departamento de Fisiología y Farmacología, Universidad de Alcalá de Henares, 28871 Madrid, Spain.

A universal characteristic of small one-domain globular protein denaturation is the "all or none" transition type, analogous to a first order phase transition for an infinite system (1). This finding may seem surprising, since proteins are irregular heterogeneous systems (2), and heterogeneity often destroy phase transitions. However, the cooperative nature of protein denaturation is of paramount importance for the protein stability. If proteins could denaturate non-cooperatively, thermal motions would destroy protein structure. Thus, proteins are unique examples of molecules with intramolecular all-or-none transition. Although there is much experimental evidence showing that unfolding is a cooperative process, there has been very little theoretical work towards explaining this phenomenon. A theory describing protein denaturation as an intramolecular melting analogous to the disruption of crystals has been developed by Shakhnovich & Filkenstein (2). In this theory, the disruption of the tightly packed side chains attached to rigid regions of secondary structure yields a first-order transition *in vacuo* , with a transition temperature of about 500 K. According with this theory, in the transition state there is no enough space to permit rotational isomerization of the side chains, while a significant part of the van der Waals energy that stabilizes the native state has already been lost. This creates unstable intermediates with high energy and low entropy, and originates the presence of the first order transition.

The experimental verification of the Shakhnovich & Filkenstein theory is complicated because of the fact that it deals with a protein *in vacuo* . Furthermore, the simplifications inherent to the theory make difficult a quantitative comparison of the predicted enthalpy and entropy of unfolding with the available experimental results. Yet, this theory is an important step in the understanding of the "all-or-none" character of protein denaturation, and it is of interest to have some measure of its range of validity. To this end, computer simulations of protein unfolding *in vacuo* can be of value. In this comunication, we report a molecular dynamics simulation of the thermal denaturation of a small globular protein *in vacuo* , and the comparison of the numerically estimated parameters describing the unfolding trajectory with

those derived form the Shakhnovich & Filkenstein theory, and with experimental data, if available. The three dimensional structure of the 434 repressor (Brookhaven Data Bank entry: 1r69), an all α-helical protein of 63 residues, was used in the simulations. Polar hydrogens were added to the model. The average screening efect of the solvent in the electrostatics interactions was model using a $\varepsilon = 4r_{ij}$, in order to compare with available experimental results. There was not truncation of the non-bonded interactions in the calculations. After energy minimization, a molecular dynamics simulation was carried out during 1 ns. The protein was heated between 0 and 1000 K at a constant rate of 1 K ps^{-1}. A time step of 2 fs was employed, and the SHAKE algorithm was used in order to remove the high frecuency motions of the hydrogen atoms. All calculations were performed with the AMBER force field using the AMBER 4.0 package (3).

During the simulation, the 434 repressor unfolds following a first order transition, as monitored following the RMS deviation from the native structure. The calculated enthalpy of unfolding *in vacuo* is about 3.2 kcal (mol of res)$^{-1}$, and it is in reasonable agreement with recent a recent experimental estimate of 2.7-3.1 kcal (mol of res)$^{-1}$ (4). The transition temperature is 550 K, thus being in qualitative agreement with the transition temperature of 500 K predicted by the Shakhnovich & Filkenstein theory. Using the two-state model for the transition, a entropy of unfolding of 5.8 cal K^{-1} (mol of res)$^{-1}$ is obtained. This figure is in reasonable agreement with previous theoretical estimates of about 5.1 cal K^{-1} (mol of res)$^{-1}$ (5). The energetic analysis of the transition shows that the van der Waals interactions are the main forces responsable for the "all-or-none" transition, while electrostatic forces make little contribution. This is consistent with the Shakhnovich & Filkenstein theory. The expansion of the protein at the transition point is about 50 %. The content of secondary structure is about 30 % of the native protein, which suggests that the assumption of the conservation of the secondary structure *in vacuo* at the transition point in the Shakhnovich & Filkenstein theory is not valid.

[1] P. L. Privalov (1979) *Adv. Protein Chem.* **33**, 167.

[2] E. I. Shakhnovich & A. V. Filkenstein (1989) *Biopolymers* **28**, 1667.

[3] D. Pearlman *et al.* (1991); AMBER 4.0; University of California, San Francisco.

[4] Y. Konishi & R. Feng (1994) *Biochemistry* **33**, 9706.

[5] J. A. Schellman (1955) *Trav. Lab. Carlsberg, Ser. Chim.* **29**, 230.

LINKING PROTEIN DYNAMICS TO RNA RECOGNITION

Lluís Ribas de Pouplana, Sunghoon Kim and Paul Schimmel
Department of Biology, Massachusetts Institute of Technology,
Cambridge, MA USA

Methionyl-tRNA Synthetase (MetRS) catalyzes the incorporation of methionine to $tRNA^{Met}$, recognizing and binding the anticodon loop of $tRNA^{Met}$ in a region centered around residue Trp 461. We have generated a library of randomly mutated enzymes between residues 439 and 468, and found that active enzymes could be obtained from sequences that contained up to 13 substitutions in that region. Simultaneously a set of inactive sequences that produce stable enzymes was obtained (S. Kim et al, PNAS **90**, 10046, 1993). The differences between the active and the inactive sequences are idiosyncratic and no pattern can be found to account for the loss of biological function.

In an attempt to explain the functional differences within the library, molecular dynamics simulations were used to study the conformational behavior of the mutated region.

Three dimensional models of MetRS with the mutant sequences from the library were built and refined with XPLOR (A. Brünger, 1992). During the dynamic simulations a large motion of the anticodon binding domain was observed in the wild type structure and the models with sequences from active enzymes. Models with inactive sequences failed to show that motion. This method can discriminate between an active and an inactive mutant differing only by a Val to Ala change at position 463. However, the modeling of this single mutation in the wild type MetRS structure did not affect the dynamics of the domain and we predicted that this mutation would be functionally neutral in the wild type context. This was confirmed experimentally.

The results indicate a role of the dynamics of MetRS on its function, perhaps contributing to the formation of the transition state through the induction of distorted tRNA conformations.

Chain folding in polymer crystals and proteins

Juan A. Subirana, Departament d'Enginyeria Química, Escola Tècnica Superior d'Enginyers Industrials de Barcelona, Universita Politècnica de Catalunya, Av. Diagonal 647, 08028 Barcelona.

Lamellar crystals with folded chains can be prepared from dilute polymer solutions. Is there any relationship of this process with protein folding? For example $(Gly)_{100}$ in solution will fold into crystals with a structure similar to the β sheet and a thickness of about 35 Å, eqivalent to 10 amino acid residues. There is a striking simmilarity with the size of globular proteins, although in the case of $(Gly)_{100}$ a lateral aggregation of folded chains is found. Due to the repetitive forces of interaction, a macroscopic crystal with constant thickness is formed, involving many polypeptide molecules. Why does polyglycine fold into a 2D-crystal instead of forming a 3D crystal? The classical explanation is that polymers fold into lamellar crystals due to kinetic reasons. However recent work in our laboratory with polyamides suggests that crystal thickness is thermodinamically controlled, i.e., each polymer folds into crystals with a thickness which corresponds to the minimum free energy of the system. It appears that the Brownian motion of the system (entropy) is concentrated on the surface folds of the crystal. In a similar way, turns in globular proteins may play a strong stabilizing role of the whole structure by concentrating the Brownian motion of the whole system. Thus turns appear as a requirement for the stability of the globular structure of proteins. Such a requirement may contribute to understand Levinthal's paradox.

Simulation of macromolecule backbone dynamics via a Flory-chain representation

John H. van Drie
Upjohn Labs
Kalamazoo, MI 49001

The dynamical behavior of applying a Maxwell-Boltzmann distribution of energies to the internal coordinates of a chain are investigated. Whereas the standard equations of molecular dynamics are expressed in Cartesian coordinates, here we investigate the dynamical equations derived from applying a Maxwell-Boltzmann distribution to the angular velocities associated with the internal coordinates of the macromolecule, the representation used by Flory in his analytic theories. This representation builds into the coordinate system the constraint that neighboring atoms are bonded together, unlike other methods which have incorporated these constraints via constrained Lagrangian dynamics (SHAKE) or by pair potentials. This approach appears to display the large-scale motions of proteins which are of greatest interest experimentally, at the expense of freezing out the smallest-scale motions (bond vibrations).

List of Invited Speakers

Workshop on

COMPUTATIONAL APPROACHES IN THE ANALYSIS AND ENGINEERING OF PROTEINS

List of Speakers

F.X. Avilés — Instituto de Biología Fundamental, Universidad Autónoma de Barcelona, 08193 Bellaterra, Barcelona (Spain).
Tel.: 34 3 581 13 15
Fax : 34 3 581 20 11

M. Billeter — HPM G42, Institut für Molekularbiologie un Biophysik, ETH-Hönggerber, CH-8093 Zürich (Switzerland).
Tel.: 41 1 633 34 87
Fax : 41 1 371 48 73

T. Blundell — ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E7HX (U.K.).
Tel.: 44 71 631 62 84
Fax : 44 71 631 68 05

J.M. Carazo — Centro Nacional de Biotecnología, CSIC, Universidad Autónoma, Cantoblanco, 28049 Madrid (Spain).
Tel.: 34 1 585 45 43
Fax : 34 1 585 45 06

C. Chothia — MRC Laboratory of Molecular Biology and Cambridge Centre of Protein Engineering, Hills Road, Cambridge CB2 2QH (U.K.).
Tel.: 44 223 40 21 00
Fax : 44 223 40 21 40

J.R. Devereux — Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI. 53711-1060 (USA).
Tel.: 1 608 231 52 00
Fax : 1 608 231 52 02

K.A. Dill — Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA. 94143 0446 (USA).
Tel.: 1 415 476 99 64
Fax : 1 415 476 06 88

D. Eisenberg — UCLA-DOE Lab of Structural Biology and Molecular Medicine, Molecular Biology Institute, and Department of Chemistry and Biochemistry, University of California, 405 Hilcard Avenue, Los Angeles, CA. 90024-1570 (USA).
Tel.: 1 310 825 37 54
Fax : 1 310 206 39 14

I. Fita	C.S.I.C. - Universidad Politécnica de Cataluña, c/Diagonal 647. 08028 Barcelona (Spain). Tel.: 34 3 401 66 85 Fax : 34 3 401 66 00
F. Gago	Departamento de Fisiología y Farmacología, Universidad de Alcalá, Campus Universitario, 28871 Alcalá de Henares, Madrid (Spain). Tel.: 34 1 885 45 14 Fax : 34 1 885 45 44
J. Greer	Department of Structural Biology, Department 46Y, Building AP9A, Abbott Laboratories, One Abbott Park Road, Abbott Park, IL. 60064-3500 (USA). Tel.: 1 708 937 69 33 Fax : 1 708 937 26 25
M. Karplus	Department of Chemistry, Harvard University, Cambridge, MA. 02138 (USA) and Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, 67000 Strasbourg (France) Fax : 33 88 41 10 20
M. Levitt	Department of Structural Biology, Stanford Medical School Stanford, CA. 94305 (USA). Tel.: 1 415 723 61 91 Fax : 1 415 723 84 64
G.M. Maggiora	Computer Aided Drug Discovery, Upjohn Laboratories, 301 Henrietta Street, Kalamazoo, MI. 49007-4940 (USA). Tel.: 1 616 384 93 50 Fax : 1 616 385 84 88
M.A. Navia	Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, MA. 02139-4211 (USA). Tel.: 1 617 576 31 11 Fax : 1 617 576 21 09
M. Orozco	Departamento de Bioquímica y Fisiología, Facultad de Química, Universidad de Barcelona, c/Marti i Franquès 1, 08028 Barcelona (Spain). Tel.: 34 3 402 12 13 Fax : 34 3 402 12 19
E. Querol	Instituto de Biología Fundamental, Universidad Autónoma de Barcelona, 08193 Bellaterra, Barcelona (Spain). Fax : 34 3 581 20 11
A. Rey	Departamento de Química Física I, Facultad de Ciencias Químicas, Universidad Complutense, Ciudad Universitaria, 28040 Madrid (Spain). Tel.: 34 1 394 41 23 Fax : 34 1 394 41 35

C. Sander — EMBL, Protein Design Group, Meyerhofstrasse 1, D-69012 Heidelberg (Germany).
Tel.: 49 6221 387 361
Fax : 49 6221 387 306

M.J. Sippl — Center for Applied Molecular Engineering, University of Salzburg TecnoZ, Jakob Haringer Str. 1, A-5020 Salzburg (Austria).
Tel.: 43 662 80 44
Fax : 43 662 45 48 89

O. Tapia — Department of Physical Chemistry, University of Uppsala, Box 532, S-751 21 Uppsala (Sweden).
Tel.: 46 18 18 25 00
Fax : 46 18 50 85 42

J. Thornton — Biochemistry and Molecular Biology Department, University College London, Gower Street, London WC1E 6BT (U.K.).
Tel.: 44 171 380 70 33
Fax : 44 171 380 71 93

A. Valencia — Centro Nacional de Biotecnología, CSIC, Universidad Autónoma, Cantoblanco, 28049 Madrid (Spain).
Tel.: 34 1 585 45 00
Fax : 34 1 585 45 06

W.F. van Gunsteren — Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETZ Zentrum, CH-8092 Zürich (Switzerland).
Tel.: 41 1 632 55 01
Fax : 41 1 632 10 39

S.J. Wodak — Unité de Conformation de Macromolécules Biologiques, Université Libre de Bruxelles, CP 160/16, Avenue Paul Héger P2, B-1050 Bruxelles (Belgium).
Tel.: 32 2 648 52 00
Fax : 32 2 648 89 54

List of Participants

Workshop on

COMPUTATIONAL APPROACHES IN THE ANALYSIS AND ENGINEERING OF PROTEINS

List of Participants

I. Alkorta — Instituto de Química Médica, Juan de la Cierva 3, 28006 Madrid (Spain).

S. Atrian — Departamento de Genética, Facultad de Biología, Universidad de Barcelona, Av. Diagonal 645, 08071 Barcelona (Spain).
Tel.: 34 3 402 14 98
Fax : 34 3 411 09 69

N.S. Boutonnet — Unité de Conformation des Macromolécules Biologiques, Université Libre de Bruxelles, CP 160/16, Av. P. Héger, 1050 Bruxelles (Belgium).
Tel.: 32 2 648 52 00
Fax : 32 2 648 89 54

G. Casari — EMBL, Meyerhofstraße 1, Heidelberg D-69117 (Germany).
Tel.: 49 6221 387 472
Fax : 49 6221 387 517

F. de la Cruz — Departamento de Biología Molecular, Universidad de Cantabria, Cardenal Herrera Oria s/n, 39011 Santander (Spain).
Tel.: 34 42 20 19 42
Fax : 34 42 20 19 45

X. Daura — Instituto de Biología Fundamental, Universidad Autónoma de Barcelona, 08193 Bellaterra Barcelona (Spain).
Tel.: 34 3 581 12 33
Fax : 34 3 581 20 11

D.R. Flower — Department of Physical Chemistry, Fisons Plc, Pharmaceuticals Division, Research & Development Labs, Bakewell Road, Loughborough, Leicestershire, LE11 ORH (U.K.).
Tel.: 44 509 611 011
Fax : 44 509 266 738

D. Frishman — European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, 69012 Heidelberg (Germany).
Tel.: 49 6221 387 231
Fax : 49 6221 387 306

M. García — Centro de Investigaciones Biológicas, CSIC, c/Velázquez 144, 28006 Madrid (Spain).
Tel.: 34 1 561 18 00 - Ext. 4334
Fax : 34 1 562 75 18

J. D. Hirst	Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA. 5213 (USA). Tel.: 1 619 554 80 73 Fax : 1 619 554 66 88
F.J. Luque	Departamento de Farmacia, Unidad de Fisicoquímica, Facultad de Farmacia, Universidad de Barcelona, Avda. Diagonal s/nº, 08028 Barcelona (Spain) Tel.: 34 3 402 45 56 Fax : 34 3 402 18 86
B.A. Luty	Laboratorium für Physikalische Chemie, ETH Zentrum, 8092 Zürich (Switzerland). Tel.: 41 1 632 68 41 Fax : 41 1 632 10 39
A.C.W. May	Department of Crystallography, Imperial Cancer Research Fund Unit of Structural Molecular Biology, Birkbeck College, University of London, Malet Street, London WC1E 7HX (U.K.). Tel.: 44 71 631 68 42 Fax : 44 71 631 68 05
R. Monsalve	Departamento de Bioquímica y Biología Molecular, Facultad de Química, Universidad Complutense, Ciudad Universitaria, 28040 Madrid (Spain). Tel.: 34 1 394 42 58 Fax : 34 1 394 41 59
B. Oliva	Instituto de Biología Fundamental, Universidad Autónoma de Barcelona,08193 Bellaterra, Barcelona (Spain).
M. Pons	Departamento de Química Orgánica, Division de C. Experimentales y Matemáticas, Universidad de Barcelona, Martí i Franqués 1-11, 08028 Barcelona (Spain). Tel.: 34 3 402 12 63 Fax : 34 3 339 78 78
L. Pugliese	Unité de Conformation des Macromolécules Biologiques, CP 160/16, Université Libre de Bruxelles, Av. P. Héger, 1050 Brussels (Belgium). Tel.: 32 2 648 52 00 Fax : 32 2 648 89 54
A. R. Ortiz	European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg (Germany). Tel.: 49 6221 387 255 Fax : 49 6221 387 517
L. Ribas de Pouplana	Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 02139 Cambridge, MA. (USA). Tel.: 1 617 253 50 39 Fax : 1 617 253 66 36
P. Sevilla	Instituto de Investigación Pluridisciplinar, Unidad de Resonancia

Magnética Nuclear, Universidad Complutense, Paseo Juan XXIII, 28040 Madrid (Spain).
Tel.: 34 1 394 32 89
Fax : 34 1 394 32 45

F. Stamato — Departamento de Química, Universidad Federal de São Carlos, Rodovia Washington Luis Km. 235, CP 676, 13560 905 São Carlos (Brazil).
Tel.: 55 162 74 82 08
Fax : 55 162 74 83 50

J.A. Subirana — Departamento de Ingeniería Química, ETSIIB, Universidad Politécnica de Cataluña, Av. Diagonal 647, 08028 Barcelona (Spain).
Tel.: 34 3 401 66 88
Fax : 34 3 401 71 50 or 66 00

J.H. Van Drie — Upjohn Research Labs., 301 Henrietta Street, Kalamazoo, MI. 49001-4940 (USA).
Tel.: 1 616 384 93 02
Fax : 1 616 385 84 88

Texts published in the
SERIE UNIVERSITARIA
by the
FUNDACIÓN JUAN MARCH
concerning workshops and courses organized within the
Plan for International Meetings on Biology (1989-1991)

246 **Workshop on Tolerance: Mechanisms and implications.**
Organized by P. Marrack and C. Martínez-A. Lectures by H. von Boehmer, J. W. Kappler, C. Martínez-A., H. Waldmann, N. Le Douarin, J. Sprent, P. Matzinger, R. H. Schwartz, M. Weigert, A. Coutinho, C. C. Goodnow, A. L. DeFranco and P. Marrack.

247 **Workshop on Pathogenesis-related Proteins in Plants.**
Organized by V. Conejero and L. C. Van Loon. Lectures by L. C. Van Loon, R. Fraser, J. F. Antoniw, M. Legrand, Y. Ohashi, F. Meins, T. Boller, V. Conejero, C. A. Ryan, D. F. Klessig, J. F. Bol, A. Leyva and F. García-Olmedo.

248 Beato, M.:
Course on DNA - Protein Interaction.

249 **Workshop on Molecular Diagnosis of Cancer.**
Organized by M. Perucho and P. García Barreno. Lectures by F. McCormick, A. Pellicer, J. L. Bos, M. Perucho, R. A. Weinberg, E. Harlow, E. R. Fearon, M. Schwab, F. W. Alt, R. Dalla Favera, P. E. Reddy, E. M. de Villiers, D. Slamon, I. B. Roninson, J. Groffen and M. Barbacid.

251 **Lecture Course on Approaches to Plant Development.**
Organized by P. Puigdoménech and T. Nelson. Lectures by I. Sussex, R. S. Poethig, M. Delseny, M. Freeling, S. C. de Vries, J. H. Rothman, J. Modolell, F. Salamini, M. A. Estelle, J. M. Martínez Zapater, A. Spena, P. J. J. Hooykaas, T. Nelson, P. Puigdoménech and M. Pagès.

252 **Curso Experimental de Electroforesis Bidimensional de Alta Resolución.**
Organizado por Juan F. Santarén. Seminarios por Julio E. Celis, James I. Garrels, Joël Vandekerckhove, Juan F. Santarén y Rosa Assiego.

253 **Workshop on Genome Expression and Pathogenesis of Plant RNA Viruses.**
Organized by F. García-Arenal and P. Palukaitis. Lectures by D. Baulcome, R. N. Beachy, G. Boccardo, J. Bol, G. Bruening, J. Burgyan, J. R. Díaz Ruiz, W. G. Dougherty, F. García-Arenal, W. L. Gerlach, A. L. Haenni, E. M. J. Jaspars, D. L. Nuss, P. Palukaitis, Y. Watanabe and M. Zaitlin.

254 **Advanced Course on Biochemistry and Genetics of Yeast.**
Organized by C. Gancedo, J. M. Gancedo, M. A. Delgado and I. L Calderón.

255 **Workshop on The Reference Points in Evolution.**
Organized by P. Alberch and G. A. Dover. Lectures by P. Alberch, P. Bateson, R. J. Britten, B. C. Clarke, S. Conway Morris, G. A. Dover, G. M. Edelman, R. Flavell, A. Fontdevila, A. García-Bellido, G. L. G. Miklos, C. Milstein, A. Moya, G. B. Müller, G. Oster, M. De Renzi, A. Seilacher, S. Stearns, E. S. Vrba, G. P Wagner, D. B. Wake and A. Wilson.

256 **Workshop on Chromatin Structure and Gene Expression.**
Organized by F. Azorín, M. Beato and A. A. Travers. Lectures by F. Azorín, M. Beato, H. Cedar, R. Chalkley, M. E. A. Churchill, D. Clark, C. Crane-Robinson, J. A. Dabán, S. C. R. Elgin, M. Grunstein, G. L. Hager, W. Hörz, T. Koller, U. K. Laemmli, E. Di Mauro, D. Rhodes, T. J. Richmond, A. Ruiz-Carrillo, R. T. Simpson, A. E. Sippel, J. M. Sogo, F. Thoma, A. A. Travers, J. Workman, O. Wrange and C. Wu.

257 **Lecture Course on Polyamines as modulators of Plant Development.**
Organized by A. W. Galston and A. F. Tiburcio. Lectures by N. Bagni, J. A. Creus, E. B. Dumbroff, H. E. Flores, A. W. Galston, J. Martin-Tanguy, D. Serafini-Fracassini, R. D. Slocum, T. A. Smith and A. F. Tiburcio.

258 **Workshop on Flower Development.**
Organized by H. Saedler, J. P. Beltrán and J. Paz Ares. Lectures by P. Albersheim, J. P. Beltrán, E. Coen, G. W. Haughn, J. Leemans, E. Lifschitz, C. Martin, J. M. Martínez-Zapater, E. M. Meyerowitz, J. Paz-Ares, H. Saedler, C. P. Scutt, H. Sommer, R. D. Thompson and K. Tran Thahn Van.

259 **Workshop on Transcription and Replication of Negative Strand RNA Viruses.**
Organized by D. Kolakofsky and J. Ortín. Lectures by A. K. Banerjee, M. A. Billeter, P. Collins, M. T. Franze-Fernández, A. J. Hay, A. Ishihama, D. Kolakofsky, R. M. Krug, J. A. Melero, S. A. Moyer, J. Ortín, P. Palese, R. G. Paterson, A. Portela, M. Schubert, D. F. Summers, N. Tordo and G. W. Wertz.

260 **Lecture Course Molecular Biology of the Rhizobium-Legume Symbiosis.**
Organized by T. Ruiz-Argüeso. Lectures by T. Bisseling, P. Boistard, J. A. Downie, D. W. Emerich, J. Kijne, J. Olivares, T. Ruiz-Argüeso, F. Sánchez and H. P. Spaink.

261 **Workshop The Regulation of Translation in Animal Virus-Infected Cells.**
Organized by N. Sonenberg and L. Carrasco. Lectures by V. Agol, R. Bablanian, L. Carrasco , M. J. Clemens, E. Ehrenfeld, D. Etchison, R. F. Garry, J. W. B. Hershey, A. G. Hovanessian, R. J. Jackson, M. G. Katze, M. B. Mathews, W. C. Merrick, D. J. Rowlands, P. Sarnow, R. J. Schneider, A. J. Shatkin, N. Sonenberg, H. O. Voorma and E. Wimmer.

263 **Lecture Course on the Polymerase Chain Reaction.**
Organized by M. Perucho and E. Martínez-Salas. Lectures by D. Gelfand, K. Hayashi, H. H. Kazazian, E. Martínez-Salas, M. Mc Clelland, K. B. Mullis, C. Oste, M. Perucho and J. Sninsky.

264 **Workshop on Yeast Transport and Energetics.**
Organized by A. Rodríguez-Navarro and R. Lagunas. Lectures by M. R. Chevallier, A. A. Eddy, Y. Eilam, G. F. Fuhrmann, A. Goffeau, M. Höfer, A. Kotyk, D. Kuschmitz, R. Lagunas, C. Leão, L. A. Okorokov, A. Peña, J. Ramos, A. Rodríguez-Navarro, W. A. Scheffers and J. M. Thevelein

265 **Workshop on Adhesion Receptors in the Immune System.**
Organized by T. A. Springer and F. Sánchez-Madrid. Lectures by S. J. Burakoff, A. L. Corbi-López, C. Figdor, B. Furie, J. C. Gutiérrez-Ramos, A. Hamann, N. Hogg, L. Lasky, R. R. Lobb, J. A. López de Castro, B. Malissen, P. Moingeon, K. Okumura, J. C. Paulson, F. Sánchez-Madrid, S. Shaw, T. A. Springer, T. F. Tedder and A. F. Williams.

266 **Workshop on Innovations on Proteases and their Inhibitors: Fundamental and Applied Aspects.**
Organized by F. X. Avilés. Lectures by T. L. Blundell, W. Bode, P. Carbonero, R. W.Carrell, C. S. Craik, T. E. Creighton, E. W. Davie, L. D. Fricker, H. Fritz, R. Huber, J. Kenny, H. Neurath, A. Puigserver, C. A. Ryan, J. J. Sánchez-Serrano, S. Shaltiel, R. L. Stevens, K. Suzuki, V. Turk, J. Vendrell and K. Wüthrich.

267 **Workshop on Role of Glycosyl-Phosphatidylinositol in Cell Signalling.**
Organized by J. M. Mato and J. Larner. Lectures by M. V. Chao, R. V. Farese, J. E. Felíu, G. N. Gaulton, H. U. Häring, C. Jacquemin, J. Larner, M. G. Low, M. Martín Lomas, J. M. Mato, E. Rodríguez-Boulan, G. Romero, G. Rougon, A. R. Saltiel, P. Strålfors and I. Varela-Nieto.

268 **Workshop on Salt Tolerance in Microorganisms and Plants: Physiological and Molecular Aspects.**
Organized by R. Serrano and J. A. Pintor-

Toro. Lectures by L. Adler, E. Blumwald, V. Conejero, W. Epstein, R. F. Gaber, P. M. Hasegawa, C. F. Higgins, C. J. Lamb, A. Läuchli, U. Lüttge, E. Padan, M. Pagès, U. Pick, J. A. Pintor-Toro, R. S. Quatrano, L. Reinhold, A. Rodríguez-Navarro, R. Serrano and R. G. Wyn Jones.

269 **Workshop on Neural Control of Movement in Vertebrates.**
Organized by R. Baker and J. M. Delgado-García. Lectures by C. Acuña, R. Baker, A. H. Bass, A. Berthoz, A. L. Bianchi, J. R. Bloedel, W. Buño, R. E. Burke, R. Caminiti, G. Cheron, J. M. Delgado-García, E. E. Fetz, R. Gallego, S. Grillner, D. Guitton, S. M. Highstein, F. Mora, F. J. Rubia Vila, Y. Shinoda, M. Steriade and P. L. Strick.

Texts published by the CENTRE FOR INTERNATIONAL MEETINGS ON BIOLOGY

1 **Workshop on What do Nociceptors Tell the Brain?**
Organized by C. Belmonte and F. Cerveró. Lectures by C. Belmonte, G. J. Bennet, J. N. Campbell, F. Cerveró, A. W. Duggan, J. Gallar, H. O. Handwerker, M. Koltzenburg, R. H. LaMotte, R. A. Meyer, J. Ochoa, E. R. Perl, H. P. Rang, P. W. Reeh, H. G. Schaible, R. F. Schmidt, J. Szolcsányi, E. Torebjörk and W. D. Willis Jr.

2 **Workshop on DNA Structure and Protein Recognition.**
Organized by A. Klug and J. A. Subirana. Lectures by F. Azorín, D. M. Crothers, R. E. Dickerson, M. D. Frank-Kamenetskii, C. W. Hilbers, R. Kaptein, D. Moras, D. Rhodes, W. Saenger, M. Salas, P. B. Sigler, L. Kohlstaedt, J. A. Subirana, D. Suck, A. Travers and J. C. Wang.

3 **Lecture Course on Palaeobiology: Preparing for the Twenty-First Century.**
Organized by F. Álvarez and S. Conway Morris. Lectures by F. Álvarez, S. Conway Morris, B. Runnegar, A. Seilacher and R. A. Spicer.

4 **Workshop on The Past and the Future of Zea Mays.**
Organized by B. Burr, L. Herrera-Estrella and P. Puigdoménech. Lectures by P. Arruda, J. L. Bennetzen, S. P. Briggs, B. Burr, J. Doebley, H. K. Dooner, M. Fromm, G. Gavazzi, C. Gigot, S. Hake, L. Herrera-Estrella, D. A. Hoisington, J. Kermicle, M. Motto, T. Nelson, G. Neuhaus, P. Puigdomenech, H. Saedler, V. Szabo and A. Viotti.

5 **Workshop on Structure of the Major Histocompatibility complex.**
Organized by A. Arnaiz-Villena and P. Parham. Lectures by A. Arnaiz-Villena, R. E. Bontrop, F. M. Brodsky, R. D. Campbell, E. J. Collins, P. Cresswell, M. Edidin, H. Erlich, L. Flaherty, F. Garrido, R. Germain, T. H. Hansen, G. J. Hämmerling, J. Klein, J. A. López de Castro, A. McMichael, P. Parham, P. Stastny, P. Travers and J. Trowsdale.

6 **Workshop on Behavioural Mechanisms in Evolutionary Perspective.**
Organized by P. Bateson and M. Gomendio. Lectures by J. R. Alberts, G. W. Barlow, P. Bateson, T. R. Birkhead, J. Carranza, C. ten Cate, F. Colmenares, N. B. Davies, R. I. M. Dunbar, J. A. Endler, M. Gomendio, T. Guilford, F. Huntingford, A. Kacelnik, J. Krebs, J. Maynard Smith, A. P. Møller, J. Moreno, G. A. Parker, T. Redondo, D. I. Rubenstein, M. J. Ryan, F. Trillmich and J. C. Wingfield.

7 **Workshop on Transcription Initiation in Prokaryotes.**
Organized by M. Salas and L. B. Rothman-Denes. Lectures by S. Adhya, H. Bujard, S. Busby, M. J. Chamberlin, R. H. Ebright, M. Espinosa, E. P. Geiduschek, R. L. Gourse, C. A. Gross, S. Kustu, J. Roberts, L. B. Rothman-Denes, M. Salas, R. Schleif, P. Stragier and W. C. Suh.

8 **Workshop on the Diversity of the Immunoglobulin Superfamily.**
Organized by A. N. Barclay and J. Vives. Lectures by A. N. Barclay, H. G. Boman, I. D. Campbell, C. Chothia, F. Díaz de Espada, I. Faye, L. García Alonso, P. R. Kolatkar, B. Malissen, C. Milstein, R. Paolini, P. Parham, R. J. Poljak, J. V. Ravetch, J. Salzer, N. E. Simister, J. Trinick, J. Vives, B. Westermark and W. Zimmermann.

9 **Workshop on Control of Gene Expression in Yeast.**
Organized by C. Gancedo and J. M. Gancedo. Lectures by T. G. Cooper, T. F. Donahue, K.-D. Entian, J. M. Gancedo, C. P. Hollenberg, S. Holmberg, W. Hörz, M. Johnston, J. Mellor, F. Messenguy, F. Moreno, B. Piña, A. Sentenac, K. Struhl, G. Thireos and R. S. Zitomer.

10 **Workshop on Engineering Plants Against Pests and Pathogens.**
Organized by G. Bruening, F. García-Olmedo and F. Ponz. Lectures by R. N. Beachy, J. F. Bol, T. Boller, G. Bruening, P. Carbonero, J. Dangl, R. de Feyter, F. García-Olmedo, L. Herrera-Estrella, V. A. Hilder, J. M. Jaynes, F. Meins, F. Ponz, J. Ryals, J. Schell, J. van Rie, P. Zabel and M. Zaitlin.

11 **Lecture Course on Conservation and Use of Genetic Resources.**
Organized by N. Jouve and M. Pérez de la Vega. Lectures by R. P. Adams, R. W. Allard, T. Benítez, J. I. Cubero, J. T. Esquinas-Alcázar, G. Fedak, B. V. Ford-Lloyd, C. Gómez-Campo, V. H. Heywood, T. Hodgkin, L. Navarro, F. Orozco, M. Pérez de la Vega, C. O. Qualset, J. W. Snape and D. Zohary.

12 **Workshop on Reverse Genetics of Negative Stranded RNA Viruses.**
Organized by G. W. Wertz and J. A. Melero. Lectures by G. M. Air, L. A. Ball, G. G. Brownlee, R. Cattaneo, P. Collins, R. W. Compans, R. M. Elliott, H.-D. Klenk, D. Kolakofsky, J. A. Melero, J. Ortín, P. Palese, R. F. Pettersson, A. Portela, C. R. Pringle, J. K. Rose and G. W. Wertz.

13 **Workshop on Approaches to Plant Hormone Action.**
Organized by J. Carbonell and R. L. Jones. Lectures by J. P. Beltrán, A. B. Bleecker, J. Carbonell, R. Fischer, D. Grierson, T. Guilfoyle, A. Jones, R. L. Jones, M. Koornneef, J. Mundy, M. Pagès, R. S. Quatrano, J. I. Schroeder, A. Spena, D. Van Der Straeten and M. A. Venis.

14 **Workshop on Frontiers of Alzheimer Disease.**
Organized by B. Frangione and J. Ávila. Lectures by J. Avila, K. Beyreuther, D. D. Cunningham, A. Delacourte, B. Frangione, C. Gajdusek, M. Goedert, Y. Ihara, K. Iqbal, K. S. Kosik, C. Milstein, D. L. Price, A. Probst, N. K. Robakis, A. D. Roses, S. S. Sisodia, C. J. Smith, W. G. Turnell and H. M. Wisniewski.

15 **Workshop on Signal Transduction by Growth Factor Receptors with Tyrosine Kinase Activity.**
Organized by J. M. Mato and A. Ullrich. Lectures by M. Barbacid, A. Bernstein, J. B. Bolen, J. Cooper, J. E. Dixon, H. U. Häring, C.- H. Heldin, H. R. Horvitz, T. Hunter, J. Martín-Pérez, D. Martín-Zanca, J. M. Mato, T. Pawson, A. Rodríguez-Tébar, J. Schlessinger, S. I. Taylor, A. Ullrich, M. D. Waterfield and M. F. White.

16 **Workshop on Intra- and Extra-Cellular Signalling in Hematopoiesis.**
Organized by E. Donnall Thomas and A. Grañena. Lectures by M. A. Brach, D. Cantrell, L. Coulombel, E. Donnall Thomas, M. Hernández-Bronchud, T. Hirano, L. H. Hoefsloot, N. N. Iscove, P. M. Lansdorp, J. J. Nemunaitis, N. A. Nicola, R. J. O'Reilly, D. Orlic, L. S. Park, R. M. Perlmutter, P. J. Quesenberry, R. D. Schreiber, J. W. Singer and B. Torok-Storb.

17 **Workshop on Cell Recognition During Neuronal Development.**
Organized by C. S. Goodman and F. Jiménez. Lectures by J. Bolz, P. Bovolenta, H. Fujisawa, C. S. Goodman, M. E. Hatten, E. Hedgecock, F. Jiménez, H. Keshishian, J. Y. Kuwada, L. Landmesser, D. D. M. O'Leary, D. Pulido, J. Raper, L. F. Reichardt, M. Schachner, M. E. Schwab, C. J. Shatz, M. Tessier-Lavigne, F. S. Walsh and J. Walter.

18 **Workshop on Molecular Mechanisms of Macrophage Activation.**
Organized by C. Nathan and A. Celada. Lectures by D. O. Adams, M. Aguet, C. Bogdan, A. Celada, J. R. David, A. Ding, R. M. Fauve, D. Gemsa, S. Gordon, J. A. M. Langermans, B. Mach, R. Maki, J. Mauël, C. Nathan, M. Röllinghoff, F. Sánchez-Madrid, C. Schindler, A. Sher and Q.-w. Xie.

19 Workshop on Viral Evasion of Host Defense Mechanisms.
Organized by M. B. Mathews and M. Esteban. Lectures by E. Domingo, L. Enjuanes, M. Esteban, B. N. Fields, B. Fleckenstein, L. Gooding, M. S. Horwitz, A. G. Hovanes-sian, M. G. Katze, I. M. Kerr, E. Kieff, M. B. Mathews, G. McFadden, B. Moss, J. Pavlovic. P. M. Pitha, J. F. Rodríguez, R. P. Ricciardi, R. H. Silverman, J. J. Skehel, G. L. Smith, P. Staeheli, A. J. van der Eb, S. Wain-Hobson, B. R. G. Williams and W. Wold.

20 Workshop on Genomic Fingerprinting.
Organized by McClelland and X. Estivill. Lectures by G. Baranton, D. E. Berg, X. Estivill, J. L. Guénet, K. Hayashi, H. S. Kwan, P. Liang, M. McClelland, J. H. Nadeau, D. L. Nelson, M. Perucho, W. Powell, J. A. Rafalski, O. A. Ryder, R. Sederoff, A. J. G. Simpson, J. Welsh and H. Zischler.

21 Workshop on DNA-Drug Interactions.
Organized by K. R. Fox and J. Portugal. Lectures by J. B. Chaires, W. A. Denny, R. E. Dickerson, K. R. Fox, F. Gago, T. Garestier, I. H. Goldberg, T. R. Krugh, J. W. Lown, L. A. Marky, S. Neidle, D. J. Patel, J. Portugal, A. Rich, L. P. G. Wakelin, M. J. Waring and C. Zimmer.

22 Workshop on Molecular Bases of Ion Channel Function.
Organized by R. W. Aldrich and J. López-Barneo. Lectures by R. W. Aldrich, C. M. Armstrong, P. Ascher, K. G. Beam, F. Bezanilla, S. C. Cannon, A. Castellano, D. Clapham, A. Ferrús, T. Hoshi, A. Konnerth, R. Latorre, J. Lerma, J. López-Barneo, R. MacKinnon, G. Mandel, A. Marty, C. Miller, B. Sakmann, B. Soria, W. Stühmer and W. N. Zagotta.

23 Workshop on Molecular Biology and Ecology of Gene Transfer and Propagation Promoted by Plasmids.
Organized by C. M. Thomas, E. M. H. Wellington, M. Espinosa and R. Díaz Orejas. Lectures by J. C. Alonso, F. Baquero, P. A. Battaglia, P. Courvalin, F. de la Cruz, D. K. Chattoraj, E. Díaz, R. Díaz Orejas, M. Espinosa, J. C. Fry, K. Gerdes, D. R. Helinski, B. Hohn, E. Lanka, V. de Lorenzo, S. Molin, K. Nordström, R. W. Pickup, C. M. Thomas, J. D. van Elsas, E. M. H. Wellington and B. Wilkins.

24 Workshop on Deterioration, Stability and Regeneration of the Brain During Normal Aging.
Organized by P. D. Coleman, F. Mora and M. Nieto-Sampedro. Lectures by J. Avila, Y.-A. Barde, P. D. Coleman, C. W. Cotman, Y. Lamour, M. P. Mattson, A. Matus, E. G. McGeer, B. S. Meldrum, J. Miquel, F. Mora, M. Nieto-Sampedro, V. H. Perry, M. P. Rathbone, G. S. Roth, R. R. Sturrock, H. B. M. Uylings and M. J. West.

25 Workshop on Genetic Recombination and Defective Interfering Particles in RNA Viruses.
Organized by J. J. Bujarski, S. Schlesinger and J. Romero. Lectures by V. Agol, P. Ahlquist, J. J. Bujarski, J. Burgyán, E. Domingo, L. Enjuanes, S. P. Goff, T. C. Hall, A. S. Huang, K. Kirkegaard, M. M. C. Lai, T. J. Morris, R. F. Ramig, D. J. Robinson, J. Romero, L. Roux, S. Schlesinger, A. E. Simon, W. J. M. Spaan, E. G. Strauss and S. Wain-Hobson.

26 Workshop on Cellular Interactions in the Early Development of the Nervous System of Drosophila.
Organized by J. Modolell and P. Simpson. Lectures by S. Artavanis-Tsakonas, J. A. Campos-Ortega, S. B. Carroll, C. Dambly-Chaudière, C. Q. Doe, R. G. Fehon, M. Freeman, A. Ghysen, V. Hartenstein, D. Hartley, Y. N. Jan, C. Klämbt, E. Knust, A. Martínez-Arias, M. Mlodzik, J. Modolell, M. A. T. Muskavitch, B.- Z. Shilo, P. Simpson, G. M. Technau, D. van Vactor and C. Zuker.

27 Workshop on Ras, Differentiation and Development
Organized by J. Downward, E. Santos and D. Martín-Zanca. Lectures by A. Balmain, G. Bollag, J. Downward, J. Erickson, L. A. Feig, F. Giráldez, F. Karim, K. Kornfeld, J. C. Lacal, A. Levitzki, D. R. Lowy, C. J. Marshall, J. Martín-Pérez, D. Martín-Zanca, J. Moscat, A. Pellicer, M. Perucho, E. Santos, J. Settleman, E. J. Taparowsky and M. Yamamoto.

28 Human and Experimental Skin Carcinogenesis
Organized by A. J. P. Klein-Szanto and M. Quintanilla. Lectures by A. Balmain, F. F.

Becker, J. L. Bos, G. T. Bowden, D. E. Brash, A. Cano, C. J. Conti, G. P. Dotto, N. E. Fusenig, J. L. Jorcano, A. J. P. Klein-Szanto, E. B. Lane, H. Nakazawa, G. Orth, M. Quintanilla, D. R. Roop, J. Schweizer, T. J. Slaga and S. H. Yuspa.

29 Workshop on the Biochemistry and Regulation of Programmed Cell Death.
Organized by J. A. Cidlowski, H. R. Horvitz, A. López-Rivas and C. Martínez-A. Lectures by J. C. Ameisen, R. Buttyan, J. A. Cidlowski, J. J. Cohen, M. K. L. Collins, T. G. Cotter, A. Eastman, D. R. Green, E. A. Harrington, T. Honjo, H. R. Horvitz, J. T. Isaacs, M. D. Jacobson, P. H. Krammer, A. López-Rivas, C. Martínez-A. S. Nagata, G. Núñez, R. W. Oppenheim, S. Orrenius, J. J. T. Owen, M. C. Raff, L. M. Schwartz and J. Yuan.

30 Workshop on Resistance to Viral Infection.
Organized by L. Enjuanes and M. M. C. Lai. Lectures by M. Ackermann, R. Blasco, L. Enjuanes, M. Esteban, P. A. Furth, F. L. Graham, L. Hennighausen, G. Keil, M. M. C. Lai, S. Makino, J. A. Melero, K. Olson, J. Ortín, E. Paoletti, C. M. Rice, H. L. Robinson, J. F. Rodríguez, B. Roizman, J. J. Rossi, P. Roy, S. G. Siddell, W. J. M. Spaan, P. Staeheli and M. Del Val.

31 Workshop on Growth and Cell Survival Factors in Vertebrate Development
Organized by M. C. Raff and F. de Pablo. Lectures by Y.-A. Barde, H. Beug, Y. Courtois, A. Efstratiadis, F. Giraldez, J. B. Gurdon, T. M. Jessell, N. Le Douarin, G. Martin, A. McMahon, D. Metcalf, E. Mitrani, F. de Pablo, P. H. Patterson, M. C. Raff, M. M. Rechler, D. B. Rifkin, J. P. Thiery and M. L. Toribio.

32 Workshop on Chromatin Structure and Gene Expression
Organized by F. Azorín, M. Beato and A. P. Wolffe. Lectures by F. Azorín, M. Beato, E. M. Bradbury, B. R. Cairns, F. J. Campoy, V. G. Corces, C. Crane-Robinson, E. Di Mauro, S. C. R. Elgin, G. Felsenfeld, W. T. Garrard, M. Grunstein, W. Hörz, J. T. Kadonaga, U. K. Laemmli, A. Ruiz-Carrillo, M. A. Surani, F. Thoma, J. O. Thomas, A. A. Travers, B. M. Turner, K. E. van Holde, A. P. Wolffe and J. L. Workman.

33 Workshop on Molecular Mechanisms of Synaptic Function
Organized by J. Lerma and P. H. Seeburg. Lectures by P. Ascher, M. V. L. Bennett, A. L. de Blas, J. Bockaert, G. L. Collingridge, M. G. Darlison, J. Garthwaite, C. E. Jahr, K. Keinänen, P. Krogsgaard-Larsen, J. Lerma, M. L. Mayer, K. Moriyoshi, R. A. Nicoll, J. P. Pin, J. Rossier, J. Sánchez-Prieto, P. H. Seeburg, Y. Stern-Bach and G. L. Westbrook.

The Centre for International Meetings on Biology
was created within the
Instituto Juan March de Estudios e Investigaciones,
a private foundation specialized in scientific activities
which complements the cultural work
of the *Fundación Juan March*.

The Centre endeavours to actively and
sistematically promote cooperation among Spanish
and foreign scientists working in the field of Biology,
through the organization of Workshops, Lecture
and Experimental Courses, Seminars,
Symposia and the Juan March Lectures on Biology.

In the last six years, a
total of 70 meetings and 6
Juan March Lecture Cycles, all
dealing with a wide range of
subjects of biological interest,
were organized within the
scope of the Centre.

Instituto Juan March de Estudios e Investigaciones
Castelló, 77 • Teléfono 34 - 1 - 435 42 40 • Fax 576 34 20
28006 Madrid (España)

The lectures summarized in this publication were presented by their authors at a workshop held on the 14th through the 16th of November, 1994, at the Instituto Juan March.

All published articles are exact reproduction of author's text.

There is a limited edition of 500 copies of this volume, available free of charge.