Thomas Lengauer, Dietmar Schomburg,
Michael S. Waterman (editors):

Molecular Bioinformatics

Dagstuhl-Seminar-Report; 46
07.09.-11.09.92 (9237)

ISSN 0940-1121

Tel.: +49-6871 - 2458
Fax: +49-6871 - 5942

Das Internationale Begegnungs- und Forschungszentrum für Informatik (IBFI) ist eine gemeinnützige GmbH. Sie veranstaltet regelmäßig wissenschaftliche Seminare, welche nach Antrag der Tagungsleiter und Begutachtung durch das wissenschaftliche Direktorium mit persönlich eingeladenen Gästen durchgeführt werden.

Bezugsadresse: Geschäftsstelle Schloß Dagstuhl
Informatik, Bau 36
Universität des Saarlandes
W - 6600 Saarbrücken
Germany
Tel.: +49 -681 - 302 - 4396
Fax: +49 -681 - 302 - 4397
e-mail: office@dag.uni-sb.de

Molecular Bioinformatics

Thomas Lengauer, Dietmar Schomburg, Michael Waterman (editors)

Dagstuhl–Seminar–Report, September 7 - 11 1992 (9237)

Workshop on Molecular Bioinformatics

Organizers:
Thomas Lengauer (GMD, Schloss Birlinghoven/University of Bonn),
Dietmar Schomburg (GBF, Braunschweig),
Michael S. Waterman (USC, Los Angeles)

September 7 - 11, 1992

Molecular Bioinformatics is a notion that one can assign to an area in applied computer science which is rapidly gaining significance. Roughly, this area is concerned with the development of methods and tools for analyzing, understanding, reasoning about and, eventually, designing large biomolecules such as DNA, RNA, and proteins with the aid of the computer.

With the knowledge in molecular biology increasing at an explosive rate, and data on genomes and their products being collected at tremendous speeds, Molecular Bioinformatics becomes an important challenge to applied computer scientists.

Before this background, the Dagstuhl Seminar on Molecular Bioinformatics brought together experts from all over the world that are working on algorithmic issues in this field. The workshop was interdisciplinary, with people from molecular biology, computer science, and applied mathematics attending. The workshop focussed on the following topics:

- Alignment of biomolecular sequences (DNA, RNA, Proteins),
- Modeling of large biomolecules, including the prediction and analysis of secondary and higher-level structure as well as spatial conformations (folding),
- Molecular dynamics and simulations of interactions between biomolecules,
- Interpreting nucleotide sequences and their role in gene regulation,
- Reading genomic sequences.

Besides the presentations, there were two organized evening discussion sessions on the topics *PAM matrices* and *Computer-Aided Drug Design*.

The experiment of bringing together researchers with widely varying backgrounds to discuss an exciting new interdisciplinary field was successful. The attendees discussed lively and often controversially, developed a sense of identity for the new field during the workshop, and went back home with new insights, problems and ideas. For the German researchers, the workshop was an ideal preparation for forming cooperations within the new funding program *Molecular Bioinformatics*

that had just been announced by the BMFT (German Ministery for Research and Technology). A few participants evaluated the workshop as their "most productive workshop experience".

We are especially grateful to the Dagstuhl office and team for their excellent organization in preparing and conducting the workshop, as well as their always

engaged and personal support in all matters. The cordial atmosphere in the *Schloß* was an essential ingredient of the success of this workshop. We are also grateful to NSF for providing a grant for financial support of intercontinental travel of the U.S. based participants of this workshop.

Program

	Monday, September 7
Lengauer	Welcome and Introductory Remarks
	Sequence Alignment
Morning Session	Chair: Gad Landau
Apostolico	Structuring Sequence Data Banks for Instantaneous Parallel Search
Manber	Approximate String Matching: Applications
Chang	Approximate Matching with Constant Fraction Error
Giegerich	Embedding Sequence Analysis in a Functional Programming Environment
Naor	Representing Suboptimal Aligments
Afternoon Session	Chair: Gaston Gonnet
Waterman	Parametric Sequence Alignment
Vingron	Interpretation of Parametric Alignment Plots
Blum	On Locally Optimal Alignments in Genetic Sequences
Manber	Approximate String Matching: Algorithms
	Tuesday, September 8
	Physical Mapping
Morning Session	Chair: Gene Lawler
Karp	The Physical Mapping Problem
Shamir	Physical Mapping of DNA and Interval Graphs
Chang	Physical Mapping in Practice
	Sequences and Trees
Gonnet	Towards the Best Possible Theory and Algorithms for Sequence Alignment
Warnow	New problems in Evolutionary Tree Construction
	Secondary and Tertiary Structure
Afternoon Session	Chair: Tim Havel
Schomburg	Computer-Aided Design of Proteins With New Properties
Sander	Protein Folding Theory
Selbig	Automatic Derivation of Patterns for Predicting Protein Structure
Smith	Prediction of Protein Domain Folding Classes, Hidden Markov Model
Evening Discussion	PAM Matrices, Gap Statistics, and Sequence Alignment Based on Dynamic Programming

	Wednesday, September 9
	Tertiary Structure
Morning Session	Chair: Chris Sander
Havel	Distance Geometry and Homology Modeling
Crippen	Prediction of the Tertiary Structure of Globular Proteins over Discrete Conformational Spaces
Chan	A Lattice Enumeration Approach to Protein Folding
	Secondary Structure
Koch	Graph Theoretical Description of Sheet Topologies
Miyano	Machine Discovery by Decision Trees over Regular Patterns
	Thursday, September 10
	Sequences
Morning Session	Chair: Udi Manber
Dress	Is Darwinism a Falsifiable Theory? Methods in Sequence Analysis
Sankoff	Climbing a Tree Through the Window
von Haeseler	Reconstructing Phylogenetic Trees
Gonnet	Text Searching Algorithms in Darwin
Afternoon Session	Chair: Temple Smith
Dress	Statistical Geometry in Sequence Space
Schmidt	Secondary Structure Problems in Coiled-coil Proteins
Zuker	RNA Secondary Structure Modeling
	Molecular Graphics
Taylor	Delegate Analysis from a Macromolecular Graphics Interface: Examples from Sequence Analysis
Evening Session	Problem Areas for Interaction Between Biochemistry and Computer Science
Havel	NMR Spectroscopy for Determination of Protein Structure
Crippen	Methods of Drug Design
	Friday, September 11
	Tertiary Structure
Morning Session	Chair: Gordon Crippen
Schulze-Kremer	Applications of Artificial Intelligence and Machine Learning to Protein Structure Analysis
Sippl	Hide and Seek on a Polyprotein
Kaden	Double-Point Chains in Proteins and Inverse Protein Folding
Schneider	Sequence-Structure Relationships in the Twighlight Zone
Lengauer	*Wrap-up*

Structuring Sequence Data Banks for Instantaneous Parallel Searching

Alberto Apostolico, University of Padua

Current molecular sequence data banks consist mainly of raw sequences with some annotation. While such a basic information will hardly be forfaited ever in the future, auxiliary data structures are being gradually introduced and studied which facilitate various kinds of searches and comparisons. For some such manipulations, serial computation is inadequate, so that efficient parallel methods are sought. We present a data structure that supports constant-time implementation on a CRCW PRAM of exact searches for arbitrary patterns into arbitrary substrings of a sequence data bank.

On the Accurate Notion of Locally Optimal Alignments and Subalignments in Genetic Sequences

Norbert Blum, University of Bonn

We review old and new results with respect to locally optimal alignments and subalignments in genetic sequences. The main theorem is that the subgraph of the edit graph, containing exactly the locally optimal subalignments, can be computed in $\Theta(nmlog(n+m))$ time, in the case that the underlying cost functions are concave.

A Lattice Enumeration Approach to Protein Folding

Hue Sun Chan, University of California at San Francisco (UCSF)

To study the protein folding problem, we use exhaustive sequence and conformational enumerations to study copolymer chains configured on lattices. These model molecules are short self-avoiding chains of hydrophobic (H) and polar (P) monomers. This simple model shows that under folding conditions, a significant fraction of H/P copolymers exhibit protein-like behavior such as high compactness, considerable amount of secondary structure, and low degeneracy of the lowest energy state. We also explore the folding kinetics of those H/P copolymers which have unique native structures. Under folding conditions, these model protein molecules collapse quickly to an ensemble of relatively compact conformations, and then re-arrange much more slowly as they seek their unique native states. Folding time of the model molecules is strongly sequence-dependent, because the arrangement of H's and P's along the sequence determines the energetic landscape of the chain's conformational space. The fastest folding sequences are those whose native structures are most accessible and least protected by energy barriers.
(This is joint work with Ken A. Dill at UCSF)

Physical Mapping in Practice

William I. Chang, Cold Spring Harbor Laboratory, N.Y. USA

Close collaboration between biologists (D. Beach Lab) and mathematical scientists (T. Marr. Lab) at C.S.H.L. resulted in the fastest mapping to date of an entire genome *S. pombe* (fission yeast, a model organism for the study of the cell cycle). Contig assembly and error analysis are done using a branch-and-bound algorithm that finds the globally optimal linear arrangement of anchors, minimizing the cost of inconsistencies in the data (false positives, false negatives, repetitive sequences). Carried out in conjunction with experiments, this analysis is used to resolve inconsistencies (by repeating experiments) and to direct further experiments. High confidence in the partially constructed map makes possible reduced laboratory work and an accelerated rate of progress.

Sublinear Approximate Matching with Constant Fraction Error

William I. Chang, Cold Spring Harbor Laboratory, N.Y. USA

Pattern matching is a classical problem of computer science, and approximate matching of sequences is motivated by molecular biology. Given a database of size n and a pattern of size m over a b-letter alphabet, we wish to find all locations in the database where the pattern occurs with at most k differences (substitutions, insertions, or deletions). There exist constants ρ_b such that for $k < \rho_b m$, k differences matching has average case complexity $\Theta((n/m)(k + log_b m))$. This algorithm requires space polynomial in the size of the pattern and can be generalized to other distance as well as similarity measures.

A Contact Potential that Recognizes the Correct Folding of Proteins

V.N. Maiorov and G.M. Crippen, University of Michigan

We have devised a continous function of interresidue contacts in globular proteins such that the X-ray crystal structure has a lower function value than that of thousands of protein-like alternative conformations. From a training set of 37 proteins and a total of 10,000 alternatives, the potential satisfies altogether 73 proteins vs. their 530,000 alternatives. In addition, another 95 highly homologous protein crystal structures are correctly treated. While the potential is intended primarily to select the native out of a large choice of rather similar or very dissimilar conformers, it can also indicate approximatlely whether the native is one of the choices.

Is Darwinism a Falsifyable Theory?

Andreas Dress, Univ. Bielefeld; H.J. Bandelt, Univ. Hamburg

Though conceived as a theory, describing and explaining what has happened rather than predicting what will happen, there is one *prediction* following from Darwin's (as well as from Lamarque's) theory which can be tested: the claim that all living species can be grouped in a sensible and consistent way into one big phylogenetic tree. Standard tree reconstruction methods - successful as they may be - are not fit to properly put this claim to a test, because all of them presuppose that what they are searching for actually is a tree structure. Hence an alternative to these methods is suggested. One looks for all decompositions $\mathcal{X} = \mathcal{A} \cup \mathcal{B}$ of the collection $\mathcal{X}$ of species under consideration into two non-empty, disjoint subsets $\mathcal{A}$ and $\mathcal{B}$ such that for any species A, A' in $\mathcal{A}$ and B, B' in $\mathcal{B}$ the *four species tree* with A next to A' and B next to B' is not the least probable of the three (non-degenerate) tree structures, one can define on our four species. One then can visualise the resulting family of decompositions, of which - by some abstract combinatorial arguments - there cannot be more than $\binom{n}{2}$ on an n-set $\mathcal{X}$, by a netted diagram which simultaneously represents - through systems of pairwise parallel edges - all found decompositions. For biological data these diagrams turn out to be almost always almost treelike - thus corroborating Darwin's ideas - while for, say, psychological data on colour similarity, the famous colour circle will be reproduced.

Statistical Geometry in Sequence Space

Andreas Dress, Univ. Bielefeld with Manfred Eigen and Ruthild Winkler Oswatitsch, MPI Göttingen

A statistical method of comparative sequence analysis that combines horizontal and vertical correlations among aligned sequences can be based on the analysis mainly of quartet combinations of sequences, considered as geometric four-point configurations in sequence space. Numerical invariants related to relative internal segment length are assigned to each such configuration and statistical averages of these invariants are established. They can be used for internal calibration of the topology of divergence and for quantitative determination of the *noise* level. Comparison with computer simulations reveals the high sensitivity of assignment of basic topologies even if much randomized. In addition, these procedures can be checked by vertical analysis of the aligned sequences to allow the study of divergencies with positionally varying substitution probabilities.

Embedding Sequence Analysis in the Functional Programming Paradigm

Robert Giegerich, University of Bielefeld

Compositionality and extensibility of analysis algorithms are important for close investigations of complex secondary structures in biosequences. Since these two properties belong to the main virtues of functional programs, a case study was performed to evaluate the viability of embedding sequence analysis in the functional paradigm. Its first part shows how an advanced pattern matching language can be implemented in a concise, transparent, and extensible way. Its second part reports on an implementation of lazy position trees, including efficiency measurements performed with three current functional language systems.

A Formal Method for the Evaluation and Comparison of a Class of Aligning Algorithms

Gaston H. Gonnet, ETH Zürich

Two aspects are considered to be the essential measures of an alignment algorithm: (a) How well does it discriminate between homologous sequences and random sequences and (b) When aligning homologous sequences, how many errors (misaligned positions) it will make. The methodology used for comparing algorithms and their associated scoring matrices, is based on simulating evolution, creating new sequences from a given one and then aligning the evolved sequence against the original one. Since we know the results of the evolution, it is directly measurable how many errors were done in the alignment. Since we can obviously produce random sequences, we can also test the discrimination of the algorithm. A couple of observations make the simulation of evolution possible. Since we restrict ourselves to the class of algorithms which are based on dynamic programming, we can simulate evolution as a Markovian process. The DP algorithms will ignore any relation between amino acids when they are compared, and will assign a cost which is constant and depends only on these amino acids. Something similar happens for insertions/deletions. The mathematical exact evaluation of the results is only possible for dynamic programming algorithms without deletions. For the complete algorithms we have to content ourselves with Monte Carlo simulation results. This is work in progress.

The Darwin System

Gaston H. Gonnet, ETH Zürich

Darwin is an interactive and programmable system for doing computations in molecular biology. *Darwin* is a descendent of the Maple system for doing computer algebra and shares its syntax and various design philosophies. *Darwin* is particularly strong in text handling and in sequence comparison. *Darwin* uses Pat indices

as an underlying structure (Pat indices are Patricia tree implementations of suffix trees) for doing: all against all sequence aligment of a database in $\Theta(N^\beta)$, $\beta < 2$; one against all sequence alignment in $\Theta(N^\alpha), \alpha < 1$; longest repetition searching: most frequent k-grams and, of course, exact matches in $\Theta(log N)$. *Darwin* also has approximate text searching (Levenshtein's distance) as a primitive. This is implemented as a partially defined DFA. In this way the algorithm adapts to the searching string and in practice is remarkable fast. The system is implemented by a kernel in C and libraries (some contributed by users) in *Darwin* itself. In runs on various Unix workstations and is distributed by e-mail at no cost. In total *Darwin* has more then 150 functions and commands, so this is clearly a very partial description of the system. We have been doing all are computations since 1990 in this system exclusively.

Reconstruction of Phylogenetic Trees

Arndt v. Haeseler, Univ. Munich

For a set $X = \{S_1, \ldots, S_n\}$ of n aligned sequences we want to reconstruct a phylogenetic tree T displaying the evolutionary relationship of the sequences. Using any dissimilarity measure $\delta : X \times X \rightarrow \mathbf{R}$, we define a neighbour relation $||_\delta$ (with respect to δ) for each quartet of sequences. We say S_1 and S_2 are neighbours with respect to S_3 and S_4 ($S_1S_2||_\delta S_3S_4$) if $\delta(S_1, S_2) + \delta(S_3, S_4) < min\{\delta(S_1, S_3) + \delta(S_2, S_4), \delta(S_1, S_4) + \delta(S_2, S_3)\}$. A similar definition of neighbourliness is made for binary unrooted trees T. We propose a method to find a tree T among all tree topologies for which the number of quartets that fulfil both relations $||_\delta$ and $||_T$ is maximal. Finally, we discuss some examples from our studies of rRNA and tRNA sequences.

Homology Modeling and Distance Geometry

Timothy F. Havel, Harvard Medical School

Distance geometry is a geometric model of molecules, wherein the structure is defined in terms of distance and chirality constraints. These are, respectively, lower and upper bounds on the interatomic distances, and the orientations of selected rigid and asymmetric tetrahedra of atoms. Distance geometry calculations are designed to reveal the geometric structure of the set of all conformations (spatial arrangements of the atoms) consistent with this information. The most important of these calculations involves computing a *conformational ensemble*, i.e. a set of conformations satisfying the constraints, but otherwise random. By analysing such an ensemble to discover new geometric properties that are uniformly present in all its members and hence are, with high probability, necessary consequences of the geometric constraints used as input, these calculations provide us with a crude but effective method of geometric reasoning. Distance geometry calculations have found numerous applications in chemistry and biology, most notably methods of determining structure from e.g. NMR data, exploring conformation space, and generating coordinates from connectivity tables. In this lecture, a new application is introduced, which obtains

the distance and chirality constraints sufficient to determine a protein structure from alignments of its sequence with homologous proteins of known structure. As an example, I have predicted the structure of the Flavodoxin from *E.coli* using as homologues the crystal structures of the Flavodoxins from *A.nidulans*, *C.beijerick*, *C.crispus* and *D.vulgaris*. The complete results of this study be found in a forthcoming issue of the journal *Molecular Simulation*.

Double Point Chains in Proteins and Inverse Protein Folding

Frieder Kaden, GMD, St. Augustin

How many possibilities are there to walk through the structure of a given protein if not only the main chain steps are allowed but also steps between residues that are far apart in the sequence but whose geometric distance is almost like that of backbone neighbours, and each residue is visited exactly once? The main chain of the protein corresponds to its ordinary sequence. Other walks through the protein lead to modified sequences that can be applied to find alignments to new protein sequences in the sense of the inverse problem of protein folding. In a suitable graph the above question appears as the NP-complete problem of finding all Hamiltonian paths. The problem is transformed into a double point problem that can be solved by a method which is a three-dimensional generalization of the formalism of L. Kauffman presented in his *Formal Knot Theory* in 1983.

Physical Mapping of Chromosomes

Richard M. Karp, University of California, Berkeley, CA, USA

We present several algorithms for reassembling the overlap structure of clones on a chromosome, given various kinds of *fingerprint data* for the clones.

Graph Theoretical Description of Sheet Topologies

Ina Koch, GMD, St. Augustin

My talk was about the usage of graph theoretical descriptions of proteins at different structure levels. We defined the protein graph, that describes the protein structure at the residue level, and the beta graph describing sheet topologies. At the protein graph level we derived patterns, which can be divided in sequentially short-range (describing helical and turn structures) and long-range patterns (describing super-secondary structures). We matched the long-range patterns against certain sheet topologies, chosen from a set of non-homologous proteins in order to find relations between patterns and topologies. First results show, that there is a quite different amino acid distribution in the patterns, which are matching against certain topologies.

Approximate Pattern Matching: New Algorithms and Applications

Udi Manber, University of Arizona

We described a tool for approximate pattern matching, called *agrep*. *Agrep* can search, very fast, for complicated patterns including arbitrary regular expressions and allowing insertions, deletions, and/or substitution errors. Examples of the use of *agrep* were shown, and two of the five new algorithms that *agrep* uses were presented. Preliminary ideas about fast approximate pattern matching in a preprocessed library of patterns, using the triangle inequality to prune the search tree, were also discussed.

Machine Discovery by Decision Trees over Regular Patterns

Satoru Miyano, Kyushu University

We describe a machine-learning system that produces hypotheses from positive and negative examples, and report some experiments on protein data using PIR and GenBank. This learning system is developed with a learning algorithm for decision trees over regular patterns, which we devised newly for this research. In the experiments on transmembrane domain identification, the system discovered very simple hypotheses with very high accuracy from a small number of positive and negative data. These hypotheses show that negative motifs, that is, motifs of negative data, play a key role in such identification. In these experiments, we classified 20 symbols of amino acid residues according to the hydropathy indices due to Kyte and Doolittle. We call such transformation of symbols an indexing. We observed that the indexing by the hydropathy indices is important in making the learning algorithm efficient and accurate. This observation inspired us with a desire to discover such an indexing itself without any help of biological knowledge but just by a learning algorithm with data. We succeeded in it by combining the above learning algorithm and the local search technique for finding indexings. We also report some experiments on signal peptides.
(This work is with S.Arikawa, S.Kuhava, S.Shimozono, A.Shimhara and T.Shimhara)

Representing Suboptimal Alignments of Biological Sequences

Dalit Naor, Stanford Univ., USA

The opytimal alignment between a pair of biological sequences that minimizes the edit-distance may not necessarily reflect the *correct* biological alignment, that is the alignment based on structure or evolutionary changes. However, in many cases the edit-distance alignment is a good approximation to the biological alignment. Suboptimal alignments are alignments whose scores lie within the neigbourhood of the

optimum, and they were suggested as alternatives to the optimal one. We study the combinatorial nature of suboptimal alignments and give a compact representation of them. We define a *canonical set* of alignments, and argue that they can be viewed as the essential ones. We currently test this hypothesis with protein sequences.

Algorithmic Aspects of Protein Structure

Chris Sander, EMBL, Heidelberg

Proteins are beautiful and complicated three-dimensional structures. Their shape and biological function is coded by genetic information. There are thousands of biologically distinct protein classes. Physicist view proteins as polymer chains and try to understand the structural principles common to all. Biologists try to understand how genetic information is translated into the highly individualistic biological role of a protein. Computer scientists see proteins as graphs or space curves and try to develop algorithms that simplify the enormous combinatorial complexity of picking out the correct structure for a given genetic sequence.
(During the workshop we proved that the protein folding problem is SO-hard, U. Manber and C. Sander, unpublished)

Climbing a Tree Through the Window

David Sankoff, University of Montreal

The method of nearest-neighbour interchange effects local improvements in a binary tree by replacing a 4-subtree by one of its two alternatives if this improves the objective function. We extend this to k-subtrees in order to reduce the number of local optima. Possible sequences of k-subtrees to be examined are produced by moving a window over the tree, incorporating one edge at a time while deactivating another. The direction of this movement is chosen according to a hill-climbing strategy. The algorithm includes a backtracking component. Series of simulations of molecular evolution data/parsimony analysis are carried out, for $k = 4, \ldots, 8$, contrasting the hill-climbing strategy to one based on a random choice of next window, and comparing two stopping rules. Increasing window size k is found to be the most effective way of improving the local optimum, followed by the choice of hill-climbing over the random strategy. A suggestion for achieving higher values of k is based on a recursive use of the hill-climbing strategy.

Secondary Structure Problems in Coiled-Coil Proteins

Jeanette P. Schmidt, Polytechnic Univ., Brooklyn, N.Y.

We describe efficient computational tools for the examination of a class of proteins that fold as alpha-helical coiled-coil proteins. In particular we detect whether a sequence of amino acids contains the 7-residue periodicity of amino acids necessary

to promote a coiled-coil conformation. A penalty matrix is used which determines the quality of the fit of a given amino acid at a given position of the heptad. A simple and efficient algorithm is presented, which can detect irregular periodic repetitions of any constant size pattern in an arbitrary text, in time that is linear in the size of the text, (where gaps are allowed and the pattern is specified by an arbitrary penalty matrix). The algorithm is used to align proteins (presumed to be coiled-coil) to the 7 positions of the heptad. The alignment is currently used to compare the outer surface of one coiled-coil protein to the other surface of a second coiled-coil protein. (Joint work with V. Fischetti, G. Landau and P. Sellers.)

Sequence-Structure Relationship in the Twighlight Zone

Reihard Schneider, EMBL, Heidelberg

The database of known protein three-dimensional structures can be significantly increased by the use of sequence homology, based on the following observations. (1) The database of known sequences, currently at more than 25000 proteins, is two orders of magnitude larger than the database of known structures. (2) The currently most powerful method of predicting protein structures is model building by homology. (3) Structural homology can be inferred from the level of sequence similarity. (4) The threshold of sequence similarity sufficient for structural homology depends strongly on the length of the alignment. This emprically derived threshold curve for structural similarity was discussed. We first quantify the relation between sequence similarity, structure similarity and alignment length by an exhaustive survey of alignments between proteins of known structure and report a homology threshold curve as a function of alignment length. We then produce a database of homology-derived secondary structure of proteins (HSSP) by aligning to each protein of known structure all sequences deemed homologous on the basis of the threshold curve. For each known protein structure, the derived database contains the aligned sequences, secondary structure, sequence variability and sequence profile. Tertiary structures of the aligned sequences are implied, but not modelled explicitly. The results are useful in assessing the structural significance of matches in sequence database searches, in deriving preferences and patterns for structure prediction, in elucidating the structural role of conserved residues and in modelling three-dimensional detail by homology. The results of a comprehensive sequence analysis of the 182 predicted open reading frames of yeast chromosome III were presented and discussed. When the results of database similarity searches are pooled with prior knowledge, a likely function can be assigned to 42% of the proteins, and a predicted 3-D structure to a third of these (14%). The function of the remaining 58% remains to be determined. An outlook for other genome projects was given. In our opinion, development in the area of protein sequence analysis should focus on three major areas. (1) improved algorithms for the detection of real, but difficult to catch, homologies and for the direct prediction of structure and function. (2) the integration of heterogenous tools into a overall working enviroment with facile exchange of information between tasks

(sequence analysis workbench) (3) direct, on-line, on-desk availability of all relevant information in the biological literature.

Computer-Aided Design of Proteins with New Properties

Dietmar Schomburg, GBF Braunschweig / CAPE

By the increase of knowledge on protein 3D-structure, the last development of molecular graphics and force field calculations, and a good understanding of structure-function correlation, a rational design of proteins with new desired properties has been possible recently. A few examples are given in the lecture. Still, the computer-based methods as sequence alignment, 3D structure prediction, and docking prediction urgently need improvement. Examples of new developments at CAPE - the German Centre of Applied Protein Engineering - were presented in the lecture.

Applications of Artificial Intelligence and Machine Learning to Protein Structures

Steffen Schulze-Kremer, Brainware GmbH, Berlin

The IPSA method aims at learning patterns of supersecondary structure from a set of known proteins. This involves selecting a list of properties of secondary structures (topological, geometrical and chemophysical); setting up a database; running learning from observation programs on that database. So far pairs of α-helices and pairs of an α-helix and a β-strand have been classified and described. Among the classes generated there are some with three secondary structures in exact agreement, although only information on two secondary structures was given, and classes that were formed completely by long range interactions. Another AI-application in Biochemistry was shown, the use of genetic algorithms. It involves a torsion angle representation using standard bond lengths and angles; the operators SELECT, MUTATE and CROSSOVER; and a very simple fitness function of the form $E = E_{tor} + E_{vdW} + E_{stat}$. Although no conformation generated resembled the native structure (of Crambin), the genetic algorithm produced very low fitness individuals. Work is going on to improve the fitness function.

Automatic Derivation of Patterns to Predict Protein Structure

Joachim Selbig, GMD, St. Augustin

Pattern-based heuristic prediction of protein structure rests upon some form of local homology where the homology information is extracted from structure data bases in a generic form by certain generalization principles. Taking into account sequentially long-range interactions requires an appropriate representation of protein structure. In particular, this holds to the case when the patterns are generated automatically

by machine learning methods. Though the pattern generation *by hand* on the base of biophysical principles is very successful machine learning methods may be used to search the available data systematically and thus to improve our understanding of the protein folding principles. Learning has been mainly viewed as inducing general concept desciptions from a learning set subdivided into classes. One of the important dimensions for characterizing learning systems is the type of the representation languages used to describe the elements of the learning set and the concepts. In our approach the data about the spatial structure of the proteins determined by X-ray crystallography are transformed into a graph description which provides the possibility to define a multitude of patterns for describing structural elements and which may be understood as discrete forms of the contact maps.

Physical Mapping of DNA and Interval Graphs

Ron Shamir, Tel Aviv University, Israel

A fundamental problem in temporal reasoning is to determine the consistency of a set of events, where for each pair of events a set of possible *atomic relations* (precedence, overlap, containment etc.) is prescribed. *Events* are assumed to be intervals on the real line. We study the problem for a simplified model, where the only atomic relations are precedence and intersection. By restricting the input to a subset of the power set of atomic relations one gets a variety of interesting combinatorial problems. We give NP-hardness results or polynomial algorithms for a variety of such restricted problems. In the DNA physical mapping problem, the chromosome corresponds to the time line and the fragments of the DNA are the events. A simplified model for the biological problem is shown to be equivalent to one of the NP-hard restrictions of the general model. Ways to exploit the many polynomial restrictions in order to expedite physical map assembly are suggested.
(Joint work with M.C. Golumbic, IBM Haifa, Israel, and in part with H. Kaplan, Tel Aviv University, Israel)

Hide and Seek on a Polyprotein

Manfred Sippl, University Salzburg

The set of experimentally determined protein structures is used to derive a knowledge based force field which in turn allows calculation of conformational energies of proteins. The final goal is the computational determination of protein structures using the knowledge based force field. The development of force fields depends on useful techniques for the assessment of the performance of the force field at each stage of development. A necessary condition is that the native fold of a protein has lowest energy compared to a number of nonnative decoys. The current version of the force field is able to identify all native folds in our data base (160 individual chains) among approximately 40,000 alternatives. At the current state of development the force field can be used to validate experimentally determined structures, to detect

native like folds for sequences of unknown structure in a data base of known folds using sequence structure alignment techniques and to build models of entire folds from ensembles of overlapping fragments.

The Prediction of Some Protein Structural Features

Temple F. Smith, Boston University

Using an alignment of some 127 proteins from the PDB with their *close* homologs a number of statistical measure were obtained: these included the conditional probabilities:
$P(S_k|S_{k-1}), P(S_k|a_k), \ldots, P(S_k|a_k, a_{k-1}, S_{k-1})$. From these the missing or *Shannon* information was calculated, given that there are only eight structural states. These data suggest that the standard secondary structure prediction can do no better than 65% which is inaccord with experience. In addition these data suggest that the use of the variability at aligned homologous positions provides the largest reduction in missing information with an upper limit of 86% on predictions fully exploiting such information. Finally if secondary structure prediction is done in the context of our understandings of realizable tertiary structures much higher values may be possible. This was tested by modeling the tertiary 3-D information of a set of seven domain classes using discrete space state models. These models condition the primary and secondary structure on allowed tertiary structures and appear to increase the predictability to near 95%.

Interpretation of Parametric Alignment Plots

Martin Vingron, Univ. of Southern California

Based on the methods presented by M. Waterman (see above) the patterns arising for parametric alignments were analyzed. The logic behind the plot was first exemplified based on the comparison of random sequences. The most striking feature there is the clear reflection of the statistical features of alignment score in the combinatorial structure for the plots. When moving to the comparison of real biological sequences these features can be found again in a somewhat distorted form though. Nevertheless, we could highlight certain patterns which seem linked to biologically correct alignments and which might aid in their recognition.

New Problems in Evolutionary Tree Construction

Tandy J. Warnow, Sandia National Labs, USA

Classical models for constructing trees from discrete data either use distance matrices or qualitative characters. The complexity of these problems are discussed, and the flaws in these models are examined. Several new models for tree construction are

then presented which may permit efficient algorithms to be discovered, and which avoid some of the inherent limitations of the classical formulations.

Parametric Sequence Alignment

Michael Waterman, Univ. Southern California

Dynamic programming algorithms for optimal alignments of two nucleic acid or protein sequences require setting penalty parameters. While the choice of these parameters greatly influences the quality of the resulting alignments, this choice has been made in an *ad hoc* fashion. In this talk we present an algorithm to find optimal alignment scores for all choices of the penalty parameters when the score is linear in the penalty parameters. In addition some statistical theory of the asymptotic growth of alignment scores with the length of random sequences is presented and related to the parametric sequence alignments.

RNA Secondary Structure Modeling

Michael Zuker, NRC, Ottawa, Canada

RNA secondary structure modeling differs fundamentally from conventional atomic resolution modeling. By borrowing discrete optimization methods used in sequence alignment, it can unfailingly predict minimum free energy as well as close to optimal foldings. The definition of secondary structure and the recursion for computing optimal foldings are given. The dynamic programming fill algorithm runs in the time $\Theta(n^4)$ as presented, where n is the sequence size. A stopping rule is introduced that limits a backtracking step in the search for a best interior or bulge loop closed by a given base pair. It is conjectured that the expected depth of the backtracking search is bounded, resulting in an overall $\Theta(n^3)$ performance. It is shown how to predict suboptimal foldings by executing the fill algorithm on two ligated copies of the same sequence. Base pairs that can participate in optimal and close to optimal foldings are displayed as points in triangular arrays called *energy dot plots*. The energy dot plot for the entire 4217 base genome of the bacteriophage $Q\beta$ reveals distinct structural domains that correspond well to what is observed by electron microscopy. A detailed model for the central region agrees well with data from enzyme cleavage and chemical modification experiments, and is further supported by studies on mutant phages. A cluttered region in the dot plot contains base pairs of a slightly suboptimal alternate folding that is observed by electron microscopy.

Dagstuhl-Seminar 9237:

List of Participants (update: 17.09.92)

Alberto **Apostolico**
Purdue University
Computer Science Department
West Lafayette IN 47907
USA
axa@cs.purdue.edu
tel.: (317) 494 6015

Norbert **Blum**
Universität Bonn
Institut für Informatik
Römerstr. 164
W-5300 Bonn 1
blum@eos.informatik.uni-bonn.de
tel.: +49-228-550-250
fax.: +49-228-550-4 40

Hue Sun **Chan**
University of Calif. at San Francisco
Dept. of Pharmaceutical Chemistry
Box 12 04
San Francisco CA 94143
USA
chan@maxwell.mmwb.ucsf.edu
tel.: +1-415-476-89 10
fax.: +1-415-476-15 08

William I. **Chang**
Cold Spring Harbor Laboratory
100 Bungtown Road
Cold Spring Harbor NY 11724--2202
USA
wchang@cshl.org
tel.: +1-516-367-88 66
fax.: +1-516-367-84 61

Gordon **Crippen**
The University of Michigan
College of Pharmacy
Ann Arbor MI 48109-1095
USA
gordon-crippen@um.cc.umich.edu
tel.: +1-313-763-97 22

Maxime **Crochemore**
Université Paris VII
LITP
Institut Blaise Pascal
2 Place Jussieu
F-75251 Paris Cedex 05
France
mac@litp.ibp.fr
tel.: +33-1-44.27.68.47
fax.: +33-1-44.27.68.49

Andreas **Dress**
Universität Bielefeld
Fakultät Mathematik
Postfach 86 40
W-4800 Bielefeld
jordan@math25.mathematik.uni-bielefeld.de
tel.: +49-521-106-50 34/20
fax.: +49-521-106-47 43

Robert **Giegerich**
Universität Bielefeld
Technische Fakultät
Postfach 86 40
W-4800 Bielefeld
robert@techfak.uni-bielefeld.de
tel.: +49-521-106 2913

Gaston **Gonnet**
ETH Zürich
Informatik
ETH-Zentrum
CH-8092 Zürich
Switzerland
gonnet@inf.ethz.ch
tel.: +41-254-74 70
fax.: +41-1-262-39 73

Timothy **Havel**
Harvard Medical School
BCMP
240 Longwood Ave
Boston MA 02115
USA
havel@ptolemy.med.harvard.edu

Arndt **von Haeseler**
Universität München
Zoologisches Institut der Uni. München
Luisenstraße 14
W-8000 München 2
arndt@zoologie.biologie.uni-muenchen.dbp.de
tel.: +49-89-5 90 23 27
fax.: +49-89-5 90 24 74

Ralf **Hofestädt**
Universität Koblenz - Landau
Fachbereich Informatik
Rheinau 3 - 4
W-5400 Koblenz
hofestae@infko.uni-koblenz.de
tel.: +49-261-9119-431

Frieder **Kaden**
Gesellschaft für Mathematik und
Datenverarbeitung mbH
Schloß Birlinghoven
Postfach 12 40
W-5205 St. Augustin 1
Frieder.Kaden@gmd.de
tel.: +49-2241-14 27 86
fax.: +49-2241-14 21 02

Richard **Karp**
University of California at Berkeley
Computer Science Division
589 Evans Hall
Berkeley CA 94720
USA
karp@cs.berkeley.edu
tel.: +1-510-642-15 59
fax.: +1-510-642-57 75

Marek **Karpinski**
Universität Bonn
Institut für Informatik
Römerstr. 164
W-5300 Bonn 1
Germany
karpinski@cs.uni-bonn.de
tel.: +49-228-550-2 24
fax.: +49-228-550-4 40

Peter **Kleinschmidt**
Universität Passau
Fachbereich Mathematik & Informatik
Innstraße 33
W-8390 Passau
kleinsch@unipas.fmi.uni-passau.de
tel.: +49-851-50 93 39
fax.: +49-851-50 91 71

Ina **Koch**
Gesellschaft für Mathematik und
Datenverarbeitung mbH
Schloß Birlinghoven
Postfach 12 40
W-5205 St. Augustin 1
ikoch@cartan.gmd.de
tel.: +49-2241-141-28 36
fax.: +49-2241-142-8 89

Hans-Peter **Kriegel**
Universität München
Institut für Informatik
Leopoldstraße 11B
W-8000 München 40
kriegel@dbs.informatik.uni-muenchen.de
tel.: +49089-2180-52 68
fax.: +49-89-2180-52 46

Gad **Landau**
Polytechnic University
Dept. of CS
Six MetroTech Center
Brooklyn NY 11201
USA
landau@pucs2.poly.edu
tel.: +1-718-260-31 54
fax.: +1-718-260-39 06

Eugene **Lawler**
University of California at Berkeley
Computer Science Division
591 Evans Hall
Berkeley CA 94720
USA
lawler@cs.berkeley.edu
tel.: +1-510-642-40 19
fax.: +1-510-642-57 75

Thomas **Lengauer**
Gesellschaft für Mathematik und
Datenverarbeitung mbH
Schloß Birlinghoven
Postfach 12 40
W-5205 St. Augustin 1
lengauer@gmd.de
tel.: 02241-14 27 77
fax.: 02241-14 21 02

Udi **Manber**
University of Arizona
Department of Computer Science
Tucson AZ 85721
USA
udi@cs.arizona.edu
tel.: +1-602-621-66 13
fax.: +1-602-621-42 46

Saira **Mian**
University of California at Santa Cruz
Sinsheimer Laboratories
Santa Cruz CA 95064
USA
saira@fangio.UCSC.EDU
tel.: +1-408-459-27 00
fax.: +1-408-459-31 39

Satoru **Miyano**
Kyushu University 33
Research Institute of
Fundamental Information Science
Postal No. 812
Fukuoka
Japan
miyano@rifis.sci.kyushuu.ac.jp
tel.: +81-92-641-11 01 ext: 44 71
fax.: +81-92-611-26 68

Dalit **Naor**
Stanford University Medical Center
Department of Biochemistry
Stanford CA 94305-5307
USA
dnaor@cmgm.Stanford.EDU
tel.: +1-415-723-59 76
fax.: +1-415-723-67 83

Hartmut **Noltemeier**
Universität Würzburg
Lehrstuhl für Informatik
Am Hubland
W-8700 Würzburg
noltemei@uni.wuerzburg.dbp.de
tel.: +49-931-888-50 55
fax.: +49-931-888-46 00

Chris **Sander**
European Molecular Biology
Laboratory (EMBL)
Meyerhofstraße 1
W-6900 Heidelberg
sander@embl-heidelberg.de
tel.: +49-6221-38 73 61
fax.: +49-6221-38 73 06

David **Sankoff**
Université de Montreal
Centre de Recherches Mathematiques
Montreal H3C 3J7
Canada
sankoff@ere.umontreal.ca
tel.: +1-514-343-7574
fax.: +1-514-343-2254

Jeanette P. **Schmidt**
Polytechnic University
Computer Science Department
6 Metrotech Center
Brooklyn NY 11201
USA
jps@pucs4.poly.edu
tel.: +1-718-260 3502
fax.: +1-718-260-39 06

Reinhard **Schneider**
European Molecular Biology
Laboratory (EMBL)
Meyerhofstraße 1
W-6900 Heidelberg
schneider@embl-heidelberg.de
tel.: +49-6221-38 73 05
fax.: +49-6221-38 73 06

Dietmar **Schomburg**
Ges. für biotechnologische
Forschung mbH (GBF)
Mascheroder Weg 1
W-3300 Braunschweig
Germany
schomburg.@venus.gbt-braunschweig.dbp.de
tel.: +49-531-61 81-3 50
fax.: +49-531-61 81-3 55

Steffen **Schulze-Kremer**
Brainware GmbH
Gustav-Meyer-Allee 25
W-1000 Berlin 65
steffen@kristall.chemie.fu-berlin.dbp.de
tel.: +49-30-46 330-40
fax.: +49-30-469-46 49

Joachim **Selbig**
Gesellschaft für Mathematik und
Datenverarbeitung mbH
Schloß Birlinghoven
Postfach 12 40
W-5205 St. Augustin 1
selbig@cartan.gmd.de
tel.: +49-2241-14-27 92
fax.: +49-2241-14-21 02

Ron **Shamir**
Tel Aviv University
Dept. of Computer Science
Tel Aviv 69978
Israel
shamir@math.tau.ac.il
tel.: +972-3-640-93 73
fax.: +972-3-640-93 57

Manfred J. **Sippl**
Universität Salzburg
Center for Applied Molecular
Engineering
Jakob Haringer Str. 1
A-5020 Salzburg
Austria
sippl@dsbl35.edvz.sbh.ac.at
tel.: +43-662-8044-57 97

Hans-Werner **Six**
Fernuniversität-GH-Hagen
Praktische Informatik III
Feithstraße 140
W-5800 Hagen 1
six@fernuni-hagen.de
tel.: +49-2331-987-29 64
fax.: +49-2331-987-3 17

Temple **Smith**
Boston University
BMERC
36 Cummington Street
Boston MA 02215
USA
tsmith@darwin.bu.edu
tel.: +1-617-353-71 23
fax.: +1-617-353-70 20

Leslie A. **Taylor**
UCSF School of Pharmacy
Computer Graphics Lab
Box 0446 Room S926
513 Parnassus Avenue
San Francisco CA 94143-0446
USA
ltaylor@cgl.ucsf.edu
tel.: +1-415-476-5379
fax.: +1-415-502-1755

Martin **Vingron**
University of Southern California
Department of Mathematics
DRB 287 - University Park
Los Angeles CA 90089-1113
USA
mvingron@hto.usc.edu
tel.: +1-213-740-24 10
fax.: +1-213-740-24 37

Tandy **Warnow**
Princeton University
Department of Computer Science
35 Olden Street
Princeton NJ 08544
USA
twarnow@hto.usc.edu

Michael **Waterman**
University of Southern California
Department of Mathematics
DRB 287 - University Park
Los Angeles CA 90089-1113
USA
nsw@hto.usc.edu
tel.: +1-213-740-24 08

Peter **Wildmayer**
ETH Zürich
Department of Computer Science
ETH-Zentrum
CH-8092 Zürich
Switzerland
wildmayer@inf.ethz.ch
tel.: +41-1-254-74 00

Ralf **Zimmer**
Gesellschaft für Mathematik und
Datenverarbeitung mbH
Schloß Birlinghoven
Postfach 12 40
W-5205 St. Augustin 1
ralf.zimmer@gmd.de
tel.: +49-2241-14 28 18
fax.: +49-2241-14 26 18

Michael **Zuker**
National Research Council
Inst. for Biological Sciences M-54
Ottawa Ontario K1A 0R6
Canada
zuker@nrcbsz.bio.nrc.ca
tel.: +1-613-993-48 30

Zuletzt erschienene und geplante Titel:

K. Compton, J.E. Pin , W. Thomas (editors):
Automata Theory: Infinite Computations, Dagstuhl-Seminar-Report; 28, 6.-10.1.92 (9202)

H. Langmaack, E. Neuhold, M. Paul (editors):
Software Construction - Foundation and Application, Dagstuhl-Seminar-Report; 29, 13..-17.1.92 (9203)

K. Ambos-Spies, S. Homer, U. Schöning (editors):
Structure and Complexity Theory, Dagstuhl-Seminar-Report; 30, 3.-7.02.92 (9206)

B. Booß, W. Coy, J.-M. Pflüger (editors):
Limits of Modelling with Programmed Machines, Dagstuhl-Seminar-Report; 31, 10.-14.2.92 (9207)

K. Compton, J.E. Pin , W. Thomas (editors):
Automata Theory: Infinite Computations, Dagstuhl-Seminar-Report; 28, 6.-10.1.92 (9202)

H. Langmaack, E. Neuhold, M. Paul (editors):
Software Construction - Foundation and Application, Dagstuhl-Seminar-Report; 29, 13.-17.1.92 (9203)

K. Ambos-Spies, S. Homer, U. Schöning (editors):
Structure and Complexity Theory, Dagstuhl-Seminar-Report; 30, 3.-7.2.92 (9206)

B. Booß, W. Coy, J.-M. Pflüger (editors):
Limits of Information-technological Models, Dagstuhl-Seminar-Report; 31, 10.-14.2.92 (9207)

N. Habermann, W.F. Tichy (editors):
Future Directions in Software Engineering, Dagstuhl-Seminar-Report; 32; 17.2.-21.2.92 (9208)

R. Cole, E.W. Mayr, F. Meyer auf der Heide (editors):
Parallel and Distributed Algorithms; Dagstuhl-Seminar-Report; 33; 2.3.-6.3.92 (9210)

P. Klint, T. Reps, G. Snelting (editors):
Programming Environments; Dagstuhl-Seminar-Report; 34; 9.3.-13.3.92 (9211)

H.-D. Ehrich, J.A. Goguen, A. Sernadas (editors):
Foundations of Information Systems Specification and Design; Dagstuhl-Seminar-Report; 35; 16.3.-19.3.9 (9212)

W. Damm, Ch. Hankin, J. Hughes (editors):
Functional Languages:
Compiler Technology and Parallelism; Dagstuhl-Seminar-Report; 36; 23.3.-27.3.92 (9213)

Th. Beth, W. Diffie, G.J. Simmons (editors):
System Security; Dagstuhl-Seminar-Report; 37; 30.3.-3.4.92 (9214)

C.A. Ellis, M. Jarke (editors):
Distributed Cooperation in Integrated Information Systems; Dagstuhl-Seminar-Report; 38; 5.4.-9.4.92 (9215)

J. Buchmann, H. Niederreiter, A.M. Odlyzko, H.G. Zimmer (editors):
Algorithms and Number Theory, Dagstuhl-Seminar-Report; 39; 22.06.-26.06.92 (9226)

E. Börger, Y. Gurevich, H. Kleine-Büning, M.M. Richter (editors):
Computer Science Logic, Dagstuhl-Seminar-Report; 40; 13.07.-17.07.92 (9229)

J. von zur Gathen, M. Karpinski, D. Kozen (editors):
Algebraic Complexity and Parallelism, Dagstuhl-Seminar-Report; 41; 20.07.-24.07.92 (9230)

F. Baader, J. Siekmann, W. Snyder (editors):
6th International Workshop on Unification, Dagstuhl-Seminar-Report; 42; 29.07.-31.07.92 (9231)

J.W. Davenport, F. Krückeberg, R.E. Moore, S. Rump (editors):
Symbolic, algebraic and validated numerical Computation, Dagstuhl-Seminar-Report; 43; 03.08.-07.08.92 (9232)

R. Cohen, R. Kass, C. Paris, W. Wahlster (editors):
Third International Workshop on User Modeling (UM'92), Dagstuhl-Seminar-Report; 44; 10.-13.8.92 (9233)

R. Reischuk, D. Uhlig (editors):
Complexity and Realization of Boolean Functions, Dagstuhl-Seminar-Report; 45; 24.08.-28.08.92 (9235)

Th. Lengauer, D. Schomburg, M.S. Waterman (editors):
Molecular Bioinformatics, Dagstuhl-Seminar-Report; 46; 07.09.-11.09.92 (9237)

V.R. Basili, H.D. Rombach, R.W. Selby (editors):
Experimental Software Engineering Issues, Dagstuhl-Seminar-Report; 47; 14.-18.09.92 (9238)

Y. Dittrich, H. Hastedt, P. Schefe (editors):
Computer Science and Philosophy, Dagstuhl-Seminar-Report; 48; 21.09.-25.09.92 (9239)

R.P. Daley, U. Furbach, K.P. Jantke (editors):
Analogical and Inductive Inference 1992 , Dagstuhl-Seminar-Report; 49; 05.10.-09.10.92 (9241)

E. Novak, St. Smale, J.F. Traub (editors):
Algorithms and Complexity for Continuous Problems, Dagstuhl-Seminar-Report; 50; 12.10.-16.10.92 (9242)

J. Encarnação, J. Foley (editors):
Multimedia - System Architectures and Applications, Dagstuhl-Seminar-Report; 51; 02.11.-06.11.92 (9245)

F.J. Rammig, J. Staunstrup, G. Zimmermann (editors):
Self-Timed Design, Dagstuhl-Seminar-Report; 52; 30.11.-04.12.92 (9249)

B. Courcelle, H. Ehrig, G. Rozenberg, H.J. Schneider (editors):
Graph-Transformations in Computer Science, Dagstuhl-Seminar-Report; 53; 04.01.-08.01.93 (9301)

A. Arnold, L. Priese, R. Vollmar (editors):
Automata Theory: Distributed Models, Dagstuhl-Seminar-Report; 54; 11.01.-15.01.93 (9302)

W.S. Cellary, K. Vidyasankar , G. Vossen (editors):
Versioning in Data Base Management Systems, Dagstuhl-Seminar-Report; 55; 01.02.-05.02.93 (9305)

B. Becker, R. Bryant, Ch. Meinel (editors):
Computer Aided Design and Test , Dagstuhl-Seminar-Report; 56; 15.02.-19.02.93 (9307)

M. Pinkal, R. Scha, L. Schubert (editors):
Semantic Formalisms in Natural Language Processing, Dagstuhl-Seminar-Report; 57; 23.02.-26.02.93 (9308)

H. Bibel, K. Furukawa, M. Stickel (editors):
Deduction , Dagstuhl-Seminar-Report; 58; 08.03.-12.03.93 (9310)

H. Alt, B. Chazelle, E. Welzl (editors):
Computational Geometry, Dagstuhl-Seminar-Report; 59; 22.03.-26.03.93 (9312)

J. Pustejovsky, H. Kamp (editors):
Universals in the Lexicon: At the Intersection of Lexical Semantic Theories, Dagstuhl-Seminar-Report; 60; 29.03.-02.04.93 (9313)

W. Straßer, F. Wahl (editors):
Graphics & Robotics, Dagstuhl-Seminar-Report; 61; 19.04.-22.04.93 (9316)

C. Beeri, A. Heuer, G. Saake, S.D. Urban (editors):
Formal Aspects of Object Base Dynamics , Dagstuhl-Seminar-Report; 62; 26.04.-30.04.93 (9317)