Eur Biophys J (1989) 17:167–172

European Biophysics Journal

EBJ Report

Biomolecular energy calculations using transputer technology

J. M. Goodfellow [1] and F. Vovelle [2]

[1] Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK
[2] Centre de Biophysique Moleculaire, CNRS, Orleans, France

Received May 15, 1989/Accepted May 22, 1989

Abstract. We report an application of current parallel processing transputer technology which has readily achieved a 25-fold reduction in computational time of peptide-solvent interactions.

Key words: Transputer, Occam 2, tyrosine, peptide, energy, minimisation

Potential energy calculations

Energy calculations have been used for many years to study the stability and interactions of macromolecules (Levitt 1974, 1978; Blaney et al. 1982; Wipff et al. 1983). Although the advent of supercomputers has allowed us to study the dynamics of macromolecules using computer simulation techniques, there are still applications where energy minimization may be the preferred technique. This use of *potential energy* calculations is unlikely to change until methods for calculating the more relevant *free energy* become more reliable, quicker, flexible and 'user-friendly'.

In general these studies involve the calculation of a function representing the potential energy of interaction between atoms which consists of several components relating to bond distance, bond angle and torsion angle distortions as well as electrostatic and non-bonded interactions (e.g. Weiner et al. 1986). This function is minimized with respect to intra- or intermolecular geometry. Usually the application of optimization techniques to macromolecules involves the use of gradient methods such as conjugate gradients, steepest descents or Powell's method (Powell 1964).

Such calculations can obviously be undertaken on any computer from a VAX to a CRAY supercomputer. An "average" calculation may take hours to run on a microVAX. Although molecular dynamics calculations can easily justify the use of a supercomputer, the more simple energy minimization problem may not. Moreover, the user may be more concerned with ease of use of the possibility of interactive calculations. Thus, in-house computing becomes a requisite.

Iso-energy contour maps (CARTE)

CARTE is a Fortran program written by Professor F. Vovelle (Orleans, France). The algorithm sets up a two-dimensional grid of points surrounding the solute of interest, in a plane defined by a number of solute atoms. A water molecule is placed at each grid point in turn and the energy of interaction of the water molecule with solute atoms is minimized with respect to three Euler angles which define the orientation of the water molecule in space. After the energy at each of the grid point is calculated, a contour map is generated from which the approximate positions of minima can be estimated. These approximate positions are refined using a full translational and orientational minimization using another program SITE. So far, only the code for CARTE has been parallelised as this is by far the most time consuming aspect of the calculation.

We have used this type of calculation in a number of studies including investigating the effect of different force fields (Vovelle and Goodfellow 1986) and the nature of bridging water molecule sites around "A" and "B" oligonucleotides (Vovelle et al. 1989; Vovelle and Goodfellow 1989).

Parallel processing

Several modern computers located at computer centres allow for the concurrent use of a number of processors e.g. the CRAY XMP and Intel hypercube (Hockney and Jesshope 1988). By constrast, the transputer chip is available attached (often singularly) to IBM PC microcomputers, for use in relatively small arrays and in large arrays such as through the Edinburgh concurrent computing project.

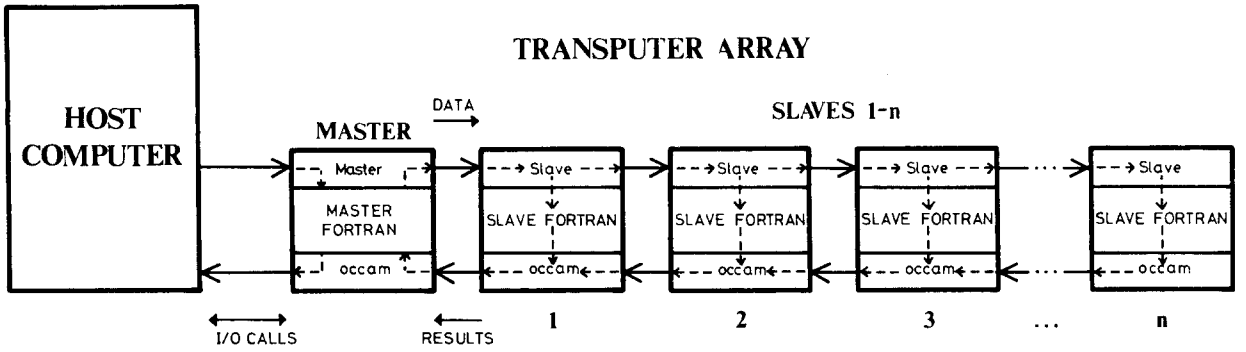

Fig. 1. Diagram of the transputer system

Our in-house configuration consists of a Meiko Computing Surface with a micro VAX front end (which is connected by Ethernet to other Vax and graphics machines). The Meiko Surface includes a host board, a master board with 8 Mbytes memory and four slave transputers with 4 Mbytes memory each (Inmos, Ltd: Transputer Reference Manual, Prentice Hall International: 1988). The host board is necessary for interaction with the front end microVAX. Thus, for any calculation, effectively five transputers are available, each of which have four channels for communication. These transputers can be arranged in any configuration. The "daisy chain" configuration is shown in Fig. 1.

Communication between the transputers has to be programmed in Occam (Inmos, Ltd: OCCAM 2 Reference Manual, Prentice Hall International 1988). Like much scientific software, the energy program, CARTE, is written in Fortran for use on conventional sequential computers. In order to deploy a Fortran program on a transputer array, it is necessary to either rewrite the code in Occam or provide an Occam harness which surrounds the Fortran code (Jones and Goldsmith 1988). We choose the latter method. The Occam harness sits on the master and slave processors and calls the Fortran routines (Fig. 1). The Fortran code on the master processors performs the initial input of parameters and data. The Fortran code on each of the slave processors is the same and performs the energy minimization calculations.

There are several strategies for writing parallel code. These include (i) dividing the Fortran code between processors so that different processors perform different parts of the calculation or (ii) dividing the data between different processors each of which performs the same calculation. One of the most important considerations is that each processor performs a similar amount of work so that one processor is not idle while another is working hard i.e. the tasks must be balanced. Moreover, one wants this balance to be achieved for a variable number of processors.

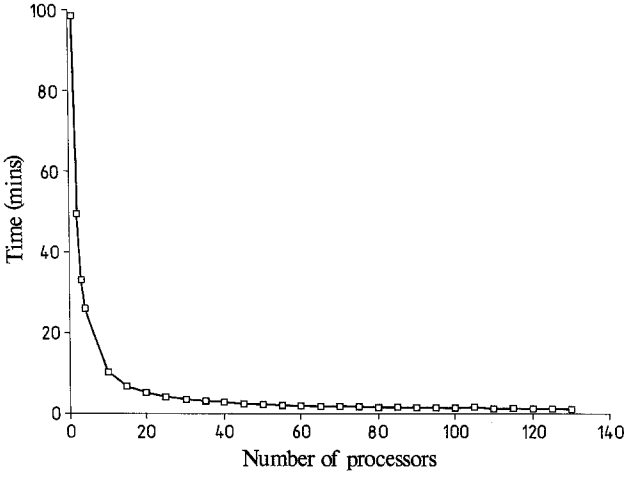

Fig. 2. Graphical representation of time for a given calculation plotted as a function of the number of processors

Initially, we chose a coarse-grained approach to parallelism in which the processors performed the same calculation but on different data points. The data points can be given to the processors in several different ways. Again, balance is important. In our application with CARTE, some of the grid points are not calculated (i.e. those points which overlap the positions of solute atoms). These all fall in one area of the map. Therefore, any division of the points into groups, which had all these null points in one group, would result in serious imbalance of work among the processors. Thus relatively well-balanced code has been achieved by giving the i^{th} transputer the i^{th} point and every following n^{th} where n is the number of slave processors.

Speed and efficiency of parallel code

The dramatic effect of running the CARTE algorithm in a parallel mode can be seen in Fig. 2. We have chosen one standard calculation for these tests. It took

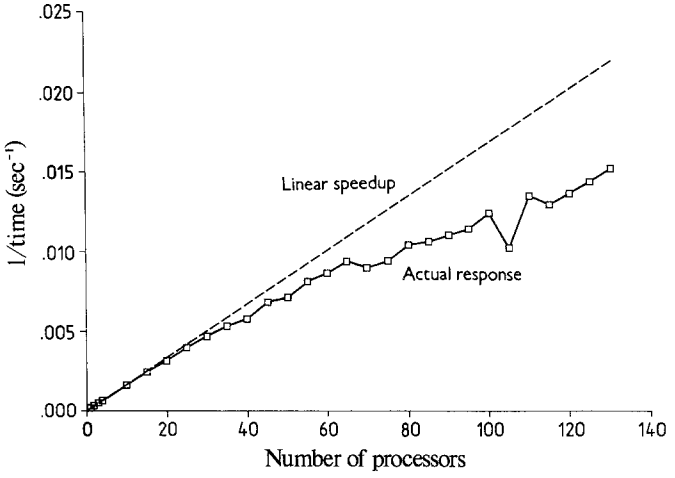

Fig. 3. Graphical representation of the speed of a given calculation plotted as a function of the number of processors

just over 10 h on a VAX 11/750. Using only four slave transputers on our in-house transputer array, the time is decreased to approximately 25 min. In order to test the efficiency of our code with varying number of processors, we used up to 131 processors on the Edinburgh concurrent computing project array of transputers. Again, we found that the time for the same standard calculation decreased to approximately 2 min.

As the number of processors is increased, the time of communications between transputers may become important. Thus, the increase in speed with number of processors is no longer linear. In Fig. 3, we have plotted speed (1/time) against number of processors in the actual and ideal (i.e. linear) case. The deviation from linearity does not seem to become significant until after 60 processors and even then the general trend of the actual speed is upwards. Distinct dips in the curve can be seen. These occur when the number of processors is 1 ½ or 2 × the number of grid points in one row.

Solvent interactions around amino acid side-chains

These energy maps have been used to study the interactions of water molecules with amino acid side-chains. This project involves two aspects: First, to use traditional energy calculations to aid in interpretation of a large amount of experimental data on solvent interactions in proteins. Second, to study electrostatic effects in side-chain interactions by comparison of the results of traditional energy calculations with similar calculations using state of the art, force fields which use distributed multipole analysis to represent the electrostatic component. Results from the first part of the project will be described in this review.

Table 1. Energy calculations on solvent interactions with a tyrosine side-chain as shown in Fig. 4a to d

	SEQ CONF	ala-tyr-ala α-helix	ala-tyr-ala β-sheet	ser-tyr-ser α-helix	lys-tyr-lys α-helix
Site 1 a–d	ENER	−10.15	−8.49	−10.07	−10.49
	HDIS	2.64	2.67	2.67	2.65
	HANG	174.00	174.00	166.00	170.00
	ATOM	$OH(H)_i$	$OH(H)_i$	$OH(H)_i$	$OH(H)_i$
Site 2 a–d	ENER	−8.73	−8.12	−8.75	−9.46
	HDIS	2.74	2.75	2.73	2.73
	HANG	179.00	177.00	178.00	177.00
	ATOM	OH_i	OH_i	OH_i	OH_i
Site 3 a–c	ENER	−7.24	−10.11	−7.98	
	HDIS	2.97	2.91	2.95	
	HANG	154.00	145.00	151.00	
	ATOM	O_{i-1}	O_{i-1}	O_{i-1}	
Site 4 a–d	ENER	−8.27	−9.86	−8.45	−11.30
	HDIS	2.81	2.83	2.71	2.85
	HANG	155.00	146.00	164.00	151.00
	ATOM	O_{i-3}	O_i	OG_{i+1}	O_{i-4}
Site 5 a–c	ENER	−7.76		−10.87	−17.16
	HDIS	2.88		2.94	2.74
	HANG	149.00		162.00	163.00
	ATOM	O_{i-3}		OG_{i+1}	NZ_{i+1}

ATOM Sequence at centre of polypeptide chain
CONF Conformation of polypeptide backbone
ENER Energy of interaction in kcal mol^{-1}
HDIS Hydrogen bond distances in Å
HANG Hydrogen bond angle in degrees
ATOM Atom and residue to which water molecule is hydrogen bonded. The central tyrosine is defined as residue i

Analysis of solvent interactions around tyrosine

In order to study the interactions of water molecules with tyrosine, we generated a polypeptide chain with defined geometry in which tyrosine occurs with six alanine residues on either side. For this, we used a program PLEGO written by Dr. M. Saqi (Birkbeck College). The iso-energy contour maps for the solvent interactions in the plane of the tyrosine ring are shown in Fig. 4a and b for alpha helical and extended chain backbone conformations. It is immediately apparent that the energy minima close to the hydroxyl (OH) group are very similar in both these cases. The contours in the lower two quadrants of the maps are clearly different but correspond to interactions with main chain atoms.

Changing the environment of the tyrosine residue by exchanging the neighbouring alanine residues with serine or lysine also causes distinct changes to the energy contours as can be seen in Fig. 4 c and d. These changes are again concentrated in the lower quadrants of the maps and are due to interactions with main chain and other side chain atoms.

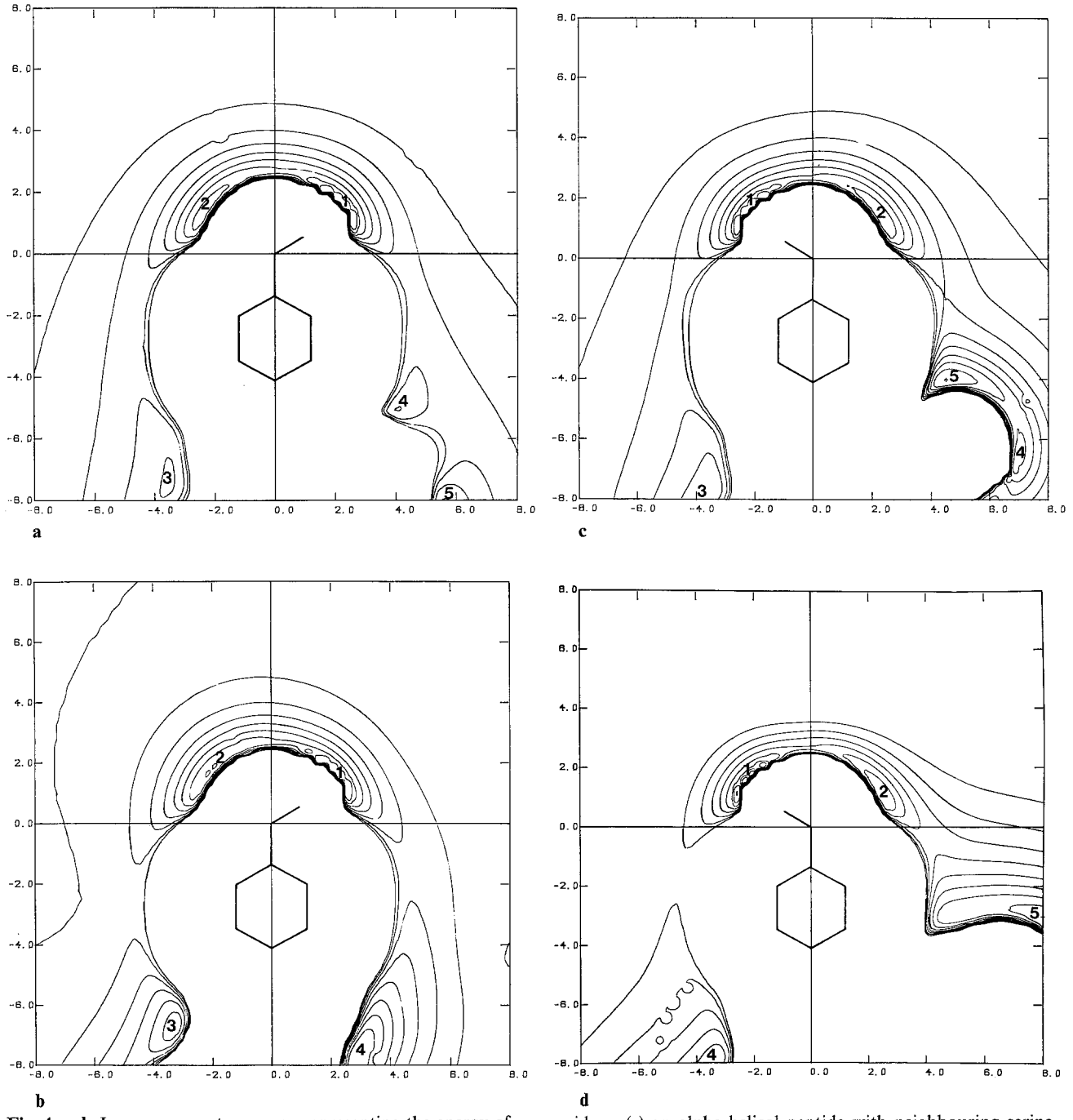

Fig. 4a–d. Iso energy contour maps representing the energy of interaction of one solvent molecule with a tyrosine side-chain in (**a**) an alpha-helical peptide with neighbouring alanine residues, (**b**) an extended chain peptide with neighbouring alanine residues, (**c**) an alpha-helical peptide with neighbouring serine residues and (**d**) an alpha-helical peptide with neighbouring lysine residues. (Distances in Å u.) The numbers of the minima refer to the sites defined in Table 1

Further analysis of these minima was undertaken using the program SITE in which the energy of interaction is minimized with respect to translational as well as orientational degrees of freedom. The results of these calculations are given in Table 1. Thus, two energy minima (sites 1a–d and sites 2a–d) are found close to the OH atom of the side-chain irrespective of the backbone conformation or type of neighbouring residues. Sites 1a–d correspond to interactions with the hydrogen of the hydroxyl group. The hydrogen bond geometry is almost identical in each case but the energy ranges from -8.49 to -10.49 kcal mol^{-1} in the extended chain and alpha helical conformation respectively. Sites 2a–d also involve hydrogen bonding interactions with hydroxyl group in which the hydrogen atom is that of the water molecule. The ener-

gies are lower than those for sites 1a–d but the hydrogen bond distances and angles are very similar.

Sites 3a–c correspond to interactions with main chain carbonyl oxygen from the residue $i-1$. It is not seen when tyrosine is surrounded by lysine residues. Sites 4a–d and 5a–c are due to interactions with main chain carbonyl atoms or polar side-chain atoms of neighbouring residues e.g. OG when residue $i+1$ is serine and N7 when residue $i+1$ is lysine. The interactions with the charged lysine NZ atom dominate the iso-energy contour map.

Discussion

The calculations were undertaken partly with the aim of comparing these energy contour maps with a detailed analysis of experimental data on the hydration of amino acid side-chains. These data were accumulated from X-ray diffraction on 16 high-resolution well refined protein structures (Thanki et al. 1988). They showed often broad distributions of solvent molecules around nearly all side-chains and distinct clustering at specific sites around most polar or charged side-chain atoms. Further analysis of the experimental data showed that some of these distributions are dependent on the local secondary structure (Goodfellow et al. 1989).

The results of the calculations on the interactions of water molecules with tyrosine are in good agreement with this experimental data. The positions of the minima, respresenting the possible positions of water molecules which are hydrogen bonded to the OH atom, are very close to the peaks in the experimental distributions. These peaks are seen in the analysis of the experimental distributions into spherical polar coordinates (r, θ, φ) centered on the OH atom of tyrosine (Thanki et al. 1988). These latter peaks occur at around $r=2.75$, $\theta=60°$ and $\varphi=0°$ or 180° (i.e. within the plane of the tyrosine ring).

The experimental solvent molecule distributions around the tyrosine OH atom showed no dependence on backbone conformation in contrast to that of the OG atom of serine and the OG1 atom of threonine. The analysis of the energy minima close to the OH atom shows no change in position with different backbone geometries or different neighbouring residues. However, these two factors do change the magnitude of the potential energy.

The energy maps also indicate the presence of other minima in or near the plane of the tyrosine ring which are due to interactions with main chain or other polar side-chain atoms. These are seen to change with the sequence of the neighbouring residues $i-1$ and $i+1$. A feature not immediately apparent from the analysis of the experimental distributions.

The computational aspects of this study are impressive. The original speed of the calculations on a conventional sequential machine is over 10 h. The use of only four slave transputers reduced the calculation by a factor of 25. Potentially greater reductions are available with increasing numbers of processors. The simple method for dividing the work load between processors is remarkably well-balanced up to 60 or more processors.

In summary, we have used the Meiko Computing Surface (a machine with highly parallel architecture) to study biomolecular interactions. Large increases in speed are obtained using Fortran code (with only minor changes) and with a simple approach to parallelism. The main difficulty in the general use of such an array of transputers lies in the writing of the Occam harness. This chore should be minimized in future by the development of more user friendly harnasses such as FORTNET being developed by Dr. R. Allan at Daresbury Laboratory.

Acknowledgements. I would like to thank Doug Jones, Roman Laskowski, David Moss, Narmada Thanki, Mansoor Saqi and Richard Westlake for their contributions to this project and the Science and Engineering Research Council for support under project grants GR/E 72737 and GR/F21183. This work was performed in part under the concurrent computing project, Edinburgh University.

References

Blaney JM, Weiner PK, Dearing A, Kollman PA, Jorgensen EC, Oatley SJ, Burridge JM, Blake CCF (1982) Molecular mechanics simulation of protein-ligand interactions: Binding of thyroid hormone analogues to prealbumin. J Am Chem Soc 104:6424–6434

Goodfellow JM, Thanki N, Thornton JM (1989) Preliminary analysis of water molecule distributions in proteins. Mol Simul 3:167–182

Hockney RW, Jesshope CR (1988) Parallel computers 2 architecture, programming and algorithms. Adam Hilger, Bristol

Jones G, Goldsmith M (1988) Programming in Occam 2. Prentice Hall, Hemel Hempstead

Levitt M (1974) Energy refinement of hen egg-white lysozyme. J Mol Biol 82:393–420

Levitt M (1978) How many base-pairs per turn does DNA have in solution and in Chromatin? Some theoretical calculation. Proc Natl Acad Sci USA 75:640–644

Powell MJD (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput J 7:155–163

Thanki N, Thornton JM, Goodfellow JM (1988) Distributions of water around amino acid residues in proteins. J Mol Biol 202:637–659

Vovelle F, Goodfellow JM (1986) The effect of different force fields on the probable hydration sites of urea. In: Sarma RH, Sarma MH (eds) Biomolecular stereodynamics III. Adenine Press, New York

Vovelle F, Goodfellow JM (1989) Modelling of DNA interactions. In: Beveridge DL, Lavery R (eds) Theoretical chemistry and molecular biophysics: A comprehensive survey. Adenine Press, New York

Vovelle F, Elliott RJ, Goodfellow JM (1989) Solvent bridging sites in A- and B-DNA helices. Int J Biol Macromol 11: 39–42

Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force field for simulation of proteins and nucleic acids. J Comp Chem 7:230–252

Wipff G, Dearing A, Weiner PK, Blaney JM, Kollman PA (1983) Molecular mechanics studies of enzyme-substrate interactions: The interaction of L- and D-N-Acetyltryptophanamide with α-chymotrypsin. J Am Chem Soc 105:997–1010

Forthcoming Meetings of Member Societies of the European Biophysical Societies' Association (EBSA)

EBSA Calendar of Events

September 17–20
1989
Todtmoos

Joint Meeting of the Austrian, German, Swiss, and Yugoslav Biophysics Societies

Topics: Structure and function of proteins, nucleic acids, and lipids (molecular biophysics); membrane biophysics and signal transduction; radiation and environmental biophysics

Information: Prof. Dr. H. Rüterjans, Institut für Biophysikalische Chemie, Theodor-Stern-Kai 7, Haus 75A, D-6000 Frankfurt/M. 70, Federal Republic of Germany

September 24–28
1989
Sevilla

Spanish Biophysical Soceity

I Congreso Iberoamericano de Biophisica

Information: I Congreso Iberoamericano de Biophisica, Departamento Fisiologia y Biofisica, Facultad de Medicina, Avenida Sanchez Pizjuan 4, E-41009 Sevilla, Spain

October 3–6
1989
Heidelberg

EMBL and Third EBSA Workshop

Patterns in protein sequences and structure

Information: Dr. Patrick Argos, EMBL, Meyerhofstrasse, D-6900 Heidelberg, Federal Republic of Germany, and Dr. Willie Taylor, National Institute for Medical Research, Mill Hill, London NW7, England