

The Art of Molecular Graphics

What Does a Molecule Look Like?

David S. Goodsell

*The Scripps Research Institute, Department of Molecular Biology,
10550 N. Torrey Pines Road, La Jolla, CA 92037 USA.
goodsell@scripps.edu, www.scripps.edu/pub/goodsell*

and

Bruce Paul Gaber

*Center for Bio/Molecular Science and Engineering, Naval Research Laboratory,
Washington, DC 20375-5348 USA. bgaber@cbmse.nrl.navy.mil*

Eighteen years ago, Mike Pique put together a short film with the title: "What Does a Protein Look Like?" The film surveyed dozens of different molecular representations that were in use and under development in Fred Brooks' laboratory at the University of North Carolina. Examples of interactive vector-based graphics, with smooth lines and perfect dots, were contrasted with raster-based images laboriously calculated one frame at a time. It was an exciting time for molecular graphics, when computer speed was just sufficient to feed an explosion of new ideas.

Today, computers are much faster and interactive molecular graphics are easily available on your desktop. Molecular graphics is no longer an experimental discipline limited to specialists, but rather a set of tested techniques used as common laboratory tools. With this transition has come a codification of representations. Three classes of molecular representation have withstood the test of time. Each may be traced to the insight of a scientist who developed a new model to reveal a new aspect of molecular matter.

At the end of the nineteenth century, G.N. Lewis revolutionized the study of chemistry by providing a set of empirical rules to describe covalent bonding. These rules are conveniently captured in a bond diagram: a line is drawn to connect each pair of atoms that are covalently bonded to one another. Bond diagrams, and their close relative, ball-and-stick diagrams, are the workhorses of chemistry and molecular biology. They omit the quantum mechanical complexity of covalent bonding, but retain its most salient features: the rigidity of bond lengths and angular geometries, and the flexibility of torsional motions. Bond diagrams became particularly popular with the advent of molecular graphics. The first systems were designed for rapidly drawing lines, so the bond diagram, already proven in chemistry, was a perfect fit for the new science of biomolecular structure.

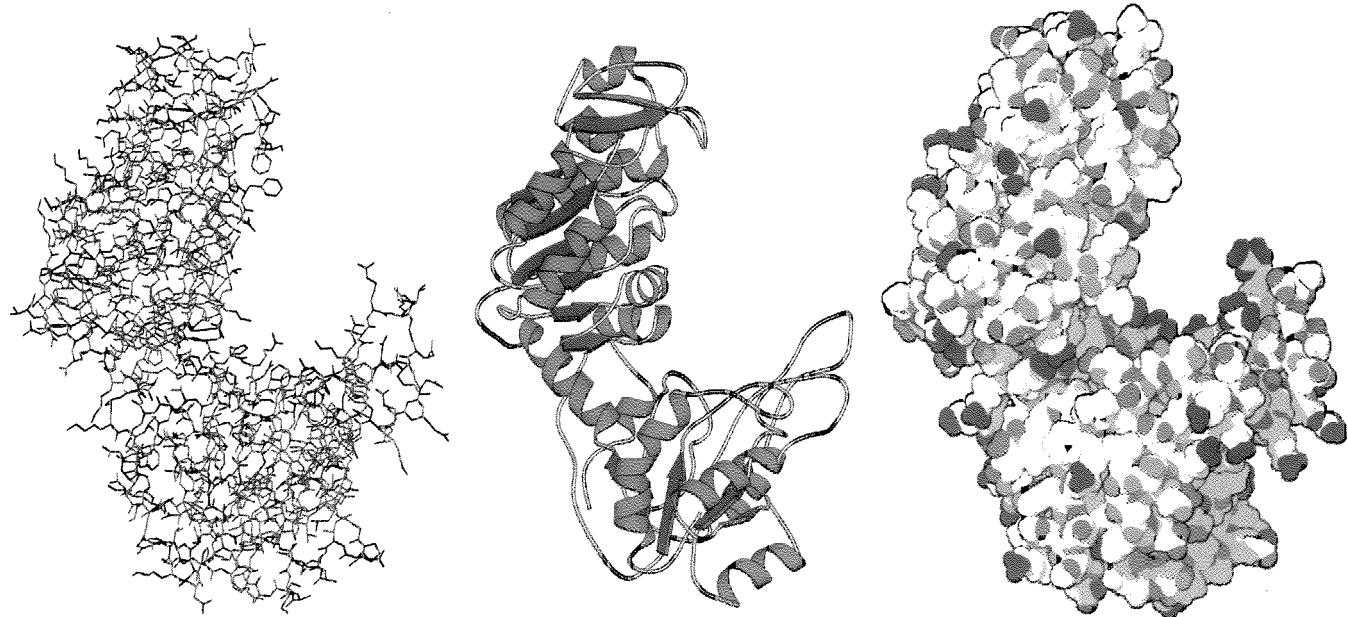
Bond diagrams, however, can be misleading. They effectively show the covalent structure of a molecule, but they discard all information on its atomistic nature — in particular, the repulsive nature of atoms at close range. Linus Pauling restored this information in a new type of representation: the spacefilling diagram. He placed a sphere around each atom, representing the inviolable space occupied by the electrons. With the availability of small plastic CPK models (named for the three scientists, Corey, Pauling, and Koltun, who perfected the manufacture of the models), spacefilling representations swept chemistry. However, they proved cumbersome when applied to biomolecular structures. Computer graphics implementation of spacefilling diagrams has progressed in fits and starts, but computers are fast enough today to provide reasonable interactivity even with large proteins. Today, spacefilling representations and their close relatives (such as solvent accessible and solvent excluded surfaces) are indispensable tools.

Bond diagrams and spacefilling diagrams, while capturing the atomic fea-

News and Views

INDEX

The Art of Molecular Graphics	219
Opinion	220
Specificity of the Tumor Necrosis Factor Receptor Superfamily	
Messages from the Chairs of COMP and MGMS	222
Meeting Reports	223
History of the Formation of the ACS Division of Computers in Chemistry	
Book Review	228
Meeting Abstracts	231
Rotational Superposition: A Review of Methods	
Teaching Computational Chemistry	238
Company Profile	244
Upcoming Meetings	247
ACS COMP Prize	248
	253



The glycolytic protein phosphoglycerate kinase (PDB accession code 3pgk) is displayed in three representations: bonds, ribbons and space-filling spheres. All are displayed at the same scale, for comparison. Notice, for instance, the large difference in the size as portrayed in ribbons and spacefilling spheres. Bonds and ribbons were calculated with MolScript (Kraulis, J. Appl. Cryst. 1991, 24, 946-950).

tures of a molecule, are very complex and often difficult to interpret. First in the study of DNA, then in the study of proteins, a schematic representation encapsulated entire structures with perfect clarity. The celebrated ladder diagram of DNA was presented by Watson and Crick in their original *Nature* article. The ladder diagram provides a clear explanation for storage and transfer of genetic information, omitting all atomic information but retaining the basic molecular concept of linear base sequence and base pairing of strands. Jane Richardson refined a similar schematic representation for proteins, creating a diagram with helical ribbons for alpha helices and arrows for beta sheets. As well as lending insight into protein folding and structure, these ribbon diagrams are quite beautiful and proved an instant success both among scientists and popularizers of science.

This is what a protein looks like at the end of the twentieth century: a collection of bonds, a cluster of spheres, or a knot of ribbons. These diagrams capture a property of the molecule, creating a representation where our intuitive understanding of the diagram matches the physical properties of the molecule. The lines capture properties of covalent bonds, the spheres express the physical bulk of atoms within molecules, and the ribbons allow us to explore chain topology. They are simple, but far from trivial, representations of molecules, imparting maximal information with a minimum of extraneous marks.

Many challenges remain for molecular graphics. Properties based on molecular interactions, such as hydrophobicity and electrostatics, are still awkward to represent. Graphical constructs from physics — isopotential surfaces and flux lines — become unmanageably complex with proteins. Coloration of surfaces seems like a good idea, until you are forced to decide between coloring by the local

potential or by projecting a potential back from a reasonable interaction distance from the surface. Neither way seems quite right. Motion and dynamics also pose real challenges. Thermal ellipsoids, coloration or sphere sizing by B-value, overlapped ensembles, and animated movie loops all have strengths and weaknesses, and all have been tried with more or less success. Of course, these properties are less intuitive than molecular bonding or chain topology, so creation of an intuitive representation is more difficult. Bond angles and geometries might be likened to the articulation of our own skeleton, but what familiar object helps us to understand an electrostatic potential?

OPINION

Specificity of the Tumor Necrosis Factor Receptor Superfamily

Jürgen Bajorath

New Chemical Entities, Inc., 18804 North Creek Pkwy. S., Bothell, WA, 98011 USA, and Department of Biological Structure, University of Washington, Seattle, WA 98195 USA.
Tel.: 425-424-7297, Fax: 425-424-7299.
jbajorath@nce-mail.com

In recent issues of *Nature Structural Biology* and *Molecular Cell*, X-ray structures of the complex formed between TRAIL (Apo2L) and one of its receptors, death receptor 5 (DR5), were reported.^{1,2} TRAIL and DR5 are transmembrane proteins belonging to the tumor necrosis factor superfamily (TNFSF) and tumor necrosis factor receptor (TNFR) superfamily (TNFRSF), respectively.³ Binding of the extracellular region of DR5 to TRAIL, a homotrimer, leads to receptor trimerization and triggers apoptosis (programmed cell death) through the intracellular death domain of DR5, similar to other TNFRSF receptors such as TNFR1 and Fas (CD95).⁴ Only two structures of TNFSF-TNFRSF ligand-receptor complexes have been reported so far, the TNF β -TNFR15 and now the TRAIL-DR5 complex. The authors have compared these structures and analyzed regions that determine binding specificity.^{1,2} On the basis of these studies, a number of general conclusions were drawn regarding the specificity of TNFRSF receptors that are a focal point of the brief analysis presented herein.

The structures of TNFRSF repeat domains in ligand-bound DR5 and TNFR1 are very similar, despite low sequence identity, but some significant differences in the relative orientation of receptor domains are observed. The ligand binding sites in DR5 and TNFR1 were analyzed in

detail to identify specificity determinants. Hymowitz et al.² have reached the conclusion that a hydrophobic segment in surface "patch B" is of general importance for ligand binding to TNFRSF receptors, whereas specificity is controlled by a segment in "patch A." These sequence segments are shown in Figure 1.

Not considered in the context of these structural investigations were previous combined molecular modeling and mutagenesis studies on CD40⁶⁻⁸ and Fas,^{9,10} other members of the TNFRSF. In these studies, single residues in CD40 and Fas were identified whose mutation significantly reduced or abolished ligand binding but not binding to panels of conformationally sensitive monoclonal antibodies.^{6,9} Residues were selected with the aid of TNFR1-based receptor models of CD40 and Fas and residues important for ligand binding were subsequently mapped using the molecular models. The results of these studies identified the ligand binding regions in CD40 and Fas and provided first evidence that non-conserved residues in spatially corresponding regions of TNFR-like receptors determine their specificity.^{7,10} These findings are now supported by structural comparison^{1,2} because the majority of previously identified residues correspond to positions in surface patches A and B of DR5 and TNFR1 (Figure 1). Moreover, residue H126

Patch B																				
TNFR1		61	T	A	S	E	N	H	L	R	H	C	*	L	S	C	S	K	* 77 C R	
DR5		51	S	T	H	W	N	D	L	L	F	C	L	R	C	T	R	C	67 D	
FAS		76										84	86	87						90 92
CD40		68	T	D	K	A	H	P	S	S	K	C	R	R	C	R	L	C	D	
												74							82 84	
			L	D	T	W	N	R	E	T	H	C	H	Q	H	K	Y	C D		

Patch A																		
TNFR1		103	Y	R	H	Y	W	S	E	N	L	F	Q	*			*	118 S
DR5		91	F	R	.	.	E	E	D	S	P	E	M					104 R
FAS		118	F	C	.	.	N	S	T	V	C	E	H	126				131 T
CD40		110	H	C	.	.	T	S	E	A	C	E	S	117				123 R

Figure 1. Ligand binding sites in TNFR-like receptors. Sequence segments forming patches A and B were aligned according to ref. 2. In X-ray structures of the TNF β -TNFR1 and TRAIL-DR5 complexes, residues in patches A and B are most intimately involved in ligand binding. Canonical cysteine positions of the TNFR fold are labeled with asterisks. Residues identified by site-specific mutagenesis as important for ligand binding to Fas and CD40 are shaded. Important residues are taken from refs. 7 and 10. Boxed residues were identified by Hymowitz et al. (ref. 2) to form the center of the binding patch in DR5 and TNFR1 and predicted to form the center in Fas and other receptors belonging to the TNFRSF.

in Fas and residue E117 in CD40 map to positions predicted to form the center of the binding site in Fas and other TNFRSF receptors² (Figure 1). These residues were previously identified as important for ligand binding.^{7,9}

The proposed importance of unique charged residues for the specificity of CD40 and Fas^{8,9} is consistent with the finding that the distribution of charged residues significantly differs in the binding sites of DR5 and TNFR1 (ref. 1). However, mutagenesis data on CD40 and Fas suggest that patch B in these receptors also contains residues important for specificity, in addition to a conserved hydrophobic motif. As shown in Figure 1, this region in CD40 and Fas contains several non-conserved and charged residues that are important for ligand recognition. Thus, the studies on CD40 and Fas provided examples for meaningful applications of molecular models in mutagenesis analyses and their results extend the conclusions drawn from structural analysis. Taken together, the independent contributions significantly improve our understanding of the specificity of the tumor necrosis factor receptor superfamily.

References

1. Mongkolsapaya, J., Grimes, J.M., Chen, N., Xu, X.-N., Stuart, D.I., Jones, E.Y., and Sreaton, G.R. Structure of the TRAIL-DR5 complex reveals mechanisms conferring specificity in apoptotic initiation. *Nature Struct. Biol.* 1999, 6:1048-1053.
2. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A., and de Vos, A.M. Triggering cell death: The crystal structure of Apo2L/TRAIL in a complex with death receptor 5. *Molecular Cell* 1999, 4:563-571.
3. Beutler, B., and van Huffel, C. Unraveling function in the TNF ligand and receptor families. *Science* 1994, 264:667-668.
4. Nagata, S. Apoptosis by death factor. *Cell* 1997, 88:355-356.
5. Banner, D.W., D'Arcy, A., Janes, W., Gentz, R., Schoenfeld, H.-J., Broger C., Loetscher, H., and Lesslauer, W. Crystal structure of the soluble human 55 kd TNF receptor-human TNF α -complex: Implications for TNF receptor activation. *Cell* 1993, 73:431-445.
6. Bajorath, J., Chalupny, N.J., Marken, J.S., Siadak, A.W., Skonier, J., Gordon, M., Hollenbaugh, D., Noelle, R.J., Ochs, H.D., and Aruffo, A. Identification of residues on CD40 and its ligand which are critical for the receptor-ligand interaction. *Biochemistry* 1995, 34:1833-1844.
7. Bajorath, J., Marken, J.S., Chalupny, N.J., Spoon, T.L., Siadak, A.W., Gordon, M., Noelle, R.J., Hollenbaugh, D., and Aruffo, A. Analysis of gp39/CD40 interactions using molecular models and site directed mutagenesis. *Biochemistry* 1995, 34:9884-9892.
8. Singh, J., Garber, E., Van Vlijmen, H., Karpusas, M., Hsu, Y.M., Zheng, Z., Naismith, J. H., and Thomas, D. The role of polar interactions in the molecular recognition of CD40L with its receptor CD40. *Protein Sci.* 1998, 7:1124-1135.
9. Starling, G.C., Bajorath, J., Emswiler, J., Ledbetter, J.A., Aruffo, A., and Kiener, P.A. Identification of amino acid residues important for ligand binding to Fas. *J. Exp. Med.* 1997, 185:1487-1492.
10. Starling, G.C., Kiener, P.A., Aruffo, A., and Bajorath, J. Analysis of the ligand binding site in Fas by site-directed mutagenesis and comparison with TNFR and CD40. *Biochemistry* 1998, 37:3723-3726.

COMP Division Chair Looks Ahead

Phil Bowen

University of Georgia, Department of Chemistry, Athens, GA 30602-2556 USA. bowen@ccmsd.chem.uga.edu

I am pleased to be able to write a few words regarding the COMP Division and its future directions. There are so many exciting events and activities being planned for next year that it is hard to begin.

First, however, I want to thank the COMP Division membership for your support by electing me as Chair-Elect (1999). Based on attending my first Executive Committee meeting in New Orleans, I can see that the position is filled with opportunities, as well as a lot of challenges. As you may know, I did not learn of the election results last year until after the formal training of 1999 ACS officers, so I am moving into this leadership role without having the benefit of formal training. Next January the new Chair-Elect (2000), David C. Spellmeyer, and I will attend the training course together in Florida. So if COMP Division news and procedures seem to start off in a bizarre or amateurish way, we can either blame it on Y2K or the lack of formal training. Second, I want to thank George Famini for his hard work and willingness to share with me what he has learned in his tenure as Chair.

Next year, the ACS national meetings are scheduled to be held in San Francisco (March 26-31) and Washington DC (August 20-25). There are a number of interesting scientific programs which have been planned, and we need to thank Dominic Ryan, Program Chair, and the Program Committee for their hard work.

The use of computers in all phases of chemistry is becoming even more widespread, which is reflected in the program for next year. I could only imagine data collection and computational chemistry when I first associated computers and chemistry. As I continue to work and develop the ACS short courses in computational chemistry and computer-assisted drug design with my co-instructors, it is apparent that there are many important applications of computers in the chemical sciences, ranging from structure calculations to 3D-database searching and library design methods. When you think about it, it is amazing how computers have become so well accepted as standard tools for chemical research and teaching.

This Division needs to continue to build on its strengths and to utilize all of its membership with all of our diverse backgrounds and interests. It is through our diversity that we can learn from each other and achieve so much. I want to challenge each of you as well as myself to help expand our membership and to participate more fully in the COMP Division programs and events in the year 2000. I hope to fulfill your expectations in my leadership style and abilities. I look forward to working with you and interacting

with you as we push the important COMP Division business to the forefront.

Molecular Graphics and Modelling Society Vision for 2000

Jonathan Essex

*Department of Chemistry, University of Southampton,
Highfield, Southampton SO17 1BJ UK
J.W.Essex@soton.ac.uk*

What is the vision for the MGMS over the coming year and beyond? Before I answer this question, it is perhaps useful to first describe the history of the Society. The Molecular Graphics Society was first constituted in 1981 "to promote and advance public education in the science of molecular graphics and related subjects, and to promote research into such subjects and make available to the public the results of such research." The Society has met this objective by holding regular scientific meetings, both in the UK and overseas, founding a Journal, and by providing financial support for students to attend the Society's meetings. In the early 1990's, it was felt that the name of the Society and its Journal did not truly reflect the breadth of science we supported. To remedy this, the Society was renamed the Molecular Graphics and Modelling Society, and the name of the Journal was similarly changed.

So how is the Society doing? Well, typically three meetings are organized each year on topics ranging from QM/MM methods to Bioinformatics, and approximately ten bursaries are made available per meeting to allow graduate students to attend and present their work. The Young Modellers' Forum is now an annual event giving graduate students the opportunity to present their work to an expert audience, and industry the chance to size-up potential employees! The Journal of Molecular Graphics and Modelling went through a lean-period several years ago, but has now been revitalized through its affiliation with the Computers in Chemistry Division of the ACS. In my view the Society is more than meeting its objectives.

Where do we go in the new millennium? It seems inevitable that the decoding of the human genome will open an enormous range of opportunities to mankind, and with that comes the challenges for our science: how do we handle the enormous amount of data, and predict protein structure and function from sequence? The Society's upcoming meeting on Structural Genomics this September will perhaps go someway to addressing these issues. For modellers, there is the prospect that the post genomics era will herald a renewed role for modelling in helping to establish the mechanisms and interactions of the cellular machinery. The increasing maturity of molecular modelling methods, coupled with the advances in computing power, make us well-placed to meet this challenge. So is it time for another addition to the Society's name? I think not yet; after all, who can say what the future will bring? However the nature of

our science may change, the MGMS will strive to meet the objectives of promoting education and research written so clearly in our constitution almost 20 years ago. Have a very happy and exciting new millennium.

MEETING REPORTS

Calculation of Concerted Motions in Biomolecules

*CECAM Workshop, Lyon, France
October 11-13, 1999*

Roger Abseher, Boehringer Ingelheim, Vienna, Austria, Michael Nilges, European Molecular Biology Laboratory, Heidelberg, Germany, István Kolossváry, Novartis Pharmaceuticals, Summit, NJ, USA, Wendy Cornell, Novartis Pharmaceuticals, Summit, NJ, USA.

In conjunction with the rapid increase in the number of three-dimensional structures available in public databases, evidence is accumulating that dynamics is a key to understanding structure-function relationships (cf. the database of macromolecular movements at <http://bioinfo.mbb.yale.edu/MolMovDB>). Concerted motions or collective degrees of freedom are increasingly being appreciated as a conceptual framework for modeling, understanding, and predicting bio-functionally important conformational diversity observed in both experiments and theoretical calculations. Current research focuses on both the definition of collective degrees of freedom and on their application in numerous fields. The CECAM (European Center for Atomic and Molecular Computations) Workshop on the "Calculation of Concerted Motions in Biomolecules" brought together approximately 30 theoreticians and computationally oriented experimentalists from 13 different countries. The basic budget provided by CECAM was supplemented by major contributions from SGI and Novartis and also by contributions from Pfizer and Pharmacia&Upjohn. The workshop was proposed and organized by Wendy Cornell, Roger Abseher, and Michael Nilges. Additional information on CECAM and its annual schedule of workshops occurring from May to October can be found at <http://www.cecam.fr>.

Collective degrees of freedom can either be determined from a single structure in a predictive manner or be calculated using multivariate statistics after some sort of sampling has taken place and generated multiple conformations. Predictive methods range from classical normal mode analysis (NMA) to a recent method put forward by I. Bahar which relaxes the requirement of putting the structure into a force field and performing extensive energy minimization. This novel method, being referred to as the Gaussian network model (GNM), proved successful in predicting relative B-factors and dominant dynamic cross-correlations for a number of biological systems where flexibility is known to be essential. Located between GNM and NMA

in terms of sophistication of the description of intramolecular interactions, there are several approximate NMA methods (K. Hinsen, Y. Sanejouand). Principal component analysis (PCA) of variance-covariance matrices (A. Amadei, D. van Aalten, B. de Groot) or singular value decomposition (SVD) of trajectory matrices (M. Pettitt) are the methods underlying approaches referred to as "essential dynamics analysis" and "quasi-normal mode analysis." The ensembles of structures serving as input for these methods may be snapshots of dynamic simulations, members of NMR structure ensembles (D. van Aalten, R. Abscher, M. Nilges), clusters of X-ray structures, or structures generated by heuristic methods (CONCOORD, B. de Groot).

Comparison of different methods reveals that there is agreement with regard to the subspace comprising the largest coordinate variances, i.e., the lowest frequency modes. However, the precise definition and ranking of modes is not transferable. A. Amadei presented measures of convergence of the essential modes derived from molecular dynamics trajectories.

The most obvious application of collective degrees of freedom is to dimensionality reduction. Typically, biological structures are soft along a limited subset of conformational coordinates. The remainder is being referred to as "constraint-like" and only marginally contributes to the overall diversity of conformational substates. L. Caves undertook first steps in the reverse direction, i.e., dimensionality expansion starting from a low-dimensional description. J. Elezgaray presented an approach which projected the (averaged) Newton equations onto a set of harmonic modes. Working in this reduced space, this method was able to fairly reproduce the average properties of the large scale motion of the protein, while using time steps one order of magnitude bigger than normal.

Upon dimensionality reduction, salient features of the conformation space topology become apparent and amenable to visualization. A tree-based method for displaying the conformation space topology has been put forward by O. Becker. N. Go's jumping-among-minima (JAM) model of conformational transitions divides fluctuations into constantly occurring harmonic ones and rare ones giving rise to anharmonicity. The time scales are correlated with barrier heights along collective coordinates.

Conformational sampling techniques making explicit use of collective degrees of freedom have become available and are being successfully applied to numerous biological problems. H. Grubmüller's conformational flooding applies a biasing potential defined along the dominant modes and thereby triggers conformational transitions. His study of the prion protein PrP lends support to Prusiner's a-to-b transition model. Also, he could rationalize a number of disease-conferring mutations using his model. I. Kolossváry demonstrated the utility of the low-mode conformational search method (LMD) for the extremely difficult problem of docking flexible ligands into flexible protein binding sites on a system that includes 9-deaza-guanine-based inhibitors docked into the flexible binding site of Purine Nucleoside Phosphorylase (PNP). The particularly successful application of LMD is attributed to the effective

use of concerted degrees of freedom of low-frequency protein motion.

D. Perahia's study of aspartate transcarbamylase (ATC) uses NMA in an attempt to understand the allosteric regulation of this complex multimeric enzyme. R. Abscher presented a method using restraints in collective coordinate space in conjunction with a standard molecular dynamics protocol (principal component restraints; PCR-MD). The successful re-folding of partially unfolded SH3 domains using PCR-MD emphasizes the role of collective re-arrangements during the late stages of protein folding.

An intriguing analysis by S. Fischer of the contribution of internal water in BPTI to the vibrational entropy revealed an increase in entropy upon binding water. D. Agard employed a normal mode based description of wild type and mutant alpha-lytic protease to define the vibrational deformations of each binding sites. These dynamic models of the active sites of the two enzymes were able to account for the specificity in binding displayed towards different inhibitors. L. Nilsson studied the dynamics of the glucocorticoid receptor DNA-binding domain where the presence of the DNA facilitates a key conformational change.

Numerous physical properties ranging from protein dielectrics (T. Simonson) to NMR dynamics measurements (R. Brueschweiler) and neutron scattering data (J. Smith) are amenable to an interpretation in the framework of collective motions. N. Go's normal mode refinement uses a smaller number of variables than does conventional X-ray refinement.

Diversity in mode space has been used as a criterion in evaluation studies of new simulation techniques, e.g., the generalized Born implicit solvent model (D. Case), and NMR structure ensembles (M. Nilges). Taking the idea of the database of macromolecular motions one step further, D. van Aalten applied the essential dynamics analysis to those systems where a sufficient number of X-ray structures are available, thereby interpreting each crystal structure as a conformational substate of the true ensemble. The highly similar subspaces found by this method and by dynamic simulations of the same systems support either approach. Individual crystal structures may be true snapshots of large-scale re-arrangements along conformational directions found in theoretical studies.

Normal mode analysis and related techniques thus provide the possibility to predict and understand concerted motions in proteins. Such a description of dynamic structure is critical to understanding many binding and signaling events currently of interest in biological systems.

Molecular Modelling '99

Bardon Conference Centre, Brisbane, Australia
22-24 September 1999

Dr. Chris Brown, School of Science, Griffith University,
Nathan, Brisbane, 4111, Queensland, Australia

The 5th Australian Molecular Modelling Conference and Workshop (MM99) was held in Brisbane from 22-24

September 1999 and was hosted by The Centre for Drug Design and Development and The University of Queensland.

This annual conference and workshop provides the main forum to foster interactions between mainly Australian researchers who have an interest in the development and uses of molecular modelling and computational chemistry. This year, our areas of interest included: chemoinformatics, scoring functions, QSAR, and molecular similarity, and attracted about 65 participants.

As in previous years, the meeting was based around lectures, short talks, posters, and "hands-on" computer sessions. The computer sessions were held at The University of Queensland's Visac lab and allowed registrants to assess the latest molecular modelling software and hardware.

The sessions were comprised of the following topics:

- bioinformatics
- chemoinformatics and molecular similarity
- de novo drug design
- scoring functions and QSAR
- protein-protein interactions
- material science

The meeting was sponsored by SGI, Tripos, Ludwig Institute for Cancer Research, Molecular Simulations Inc., GlaxoWellcome, and MDL Information Systems, Inc.

Modelling the Active Site of Acetolactate Synthase

Stephen Bowlus, Tripos Inc., San Francisco, CA, USA

Acetolactate synthase (AS) is the first enzyme in the biosynthesis of branch-chain amino acids in plants and as such is a reasonable target for contemporary herbicide research and development. Using a hybrid homology/inverse folding approach the Bowlus group has attempted to model the active site of the enzyme. *Pyruvate oxidase* was chosen as the key template as this enzyme possesses a cleft region in its active site similar to *AS*. Most of the refinement was directed in the loop region between Val 571 and Arg 583 as there was significant structural difference between the template and *AS* within this region. Subsequent site-directed mutagenesis, docking studies, and the alignment of inhibitors within the binding domain were used to rationalize the bound conformations of different herbicide classes with respect to reported resistance patterns.

Functional Genomics Using Clues from Structural Biology to Assign Gene Function

Scott D. Kahn, Molecular Simulations Inc., San Diego, CA, USA

Functional genomics is the characterization of the proteins that are expressed by a particular genome. However, given that the function of a protein is determined more directly by its overall three-dimensional structure, rather than its basic primary aa sequence, functional genomics would be vastly improved if it were possible to further process the protein sequence data derived from the genome sequence to directly generate "reasonable" 3D protein structures.

This of course would be a major asset to numerous biotechnology-based industries. Using leptin as an example, Dr. Kahn demonstrated the key steps in the structure-prediction process, employing both "in house" and shareware software coupled with a large SGI workstation.

Getting More out of Structure Databases

G. Chelvanayagam, John Curtin School of Medical Research, Australian National University, Canberra, Australia

The protein database currently comprises nearly 10,000 structure files. At present, however, most automated routines for obtaining representative datasets, for application in homology modelling studies, only select up to 1,000 structures. This falls far short of the number of unique folds expected for the human genome. Using a contemporary threading technique which scores alignments between a given sequence and a set of structures, based on statistical pair-wise potentials, Dr. Chelvanayagam is contributing to the development of multiple sequence alignment strategies — allowing homology modellers to optimize sequence fold compatibility.

Environmental Considerations of Integral Membrane Protein Structures

Siavoush Dastmalchi, Dept of Pharmacy, University of Sydney, Sydney, Australia

Integral membrane proteins (IMPs) are involved in a variety of critical cellular functions and usually compose a large fraction of expressed proteins. Whilst at present, about 25% of all available protein sequences can be modelled with a reasonable degree of success, the homology-based approach is of limited value for IMPs however, as there are few solid state structures reported for this class of protein. This lecture introduced an algorithm for the prediction of structural information for IMPs. The program takes into account the fact that many residues within the IMP reside in a local lipid environment. This environment contributes to the overall 3D structure of the protein. The Dastmalchi group have used the new algorithm to predict the location of the transmembrane segments for several IMPs.

PROFLEX: A Computational and Graphical Package for the Analysis of Movement and Flexibility in Large Molecules.

Paul A. Keller, Dept. of Chemistry, University of Wollongong, Wollongong, Australia

The Wollongong group has developed a computer program for the calculation and display of the Difference Distance Matrices of macromolecules. The PROFLEX program compares multiple structures simultaneously by selecting atoms that are invariant in all structures and is therefore freed from the normal user bias. As an example, Dr. Keller compared 21 different solid-state datasets of HIV-1 reverse transcriptase enzymes using the PROFLEX program. He found significant differences in both magnitude and trends of atom

movement when compared against other published methods for generating superimposition datasets.

Using Empiricism to Redefine Rotamer Nomenclature

Mitchell J. Polly, Swinburne University of Technology, Hawthorn, Australia

A survey of 110,000 amino acid residues from over 200 high resolution solid state protein structures revealed that classical rotamer definitions were incapable of adequately describing the behavior of some amino acid side chain dihedral angles. Using the angles collected from a survey of the solid state structures, the Swinburne group was able to provide customized rotamer definitions for several amino acid residue types. The new library provides an important extension to existing amino acid side-chain geometry data.

A Keener Understanding of Biological Problems Using Molecular Modelling

Shoba Ranganathan, Australian Genomic Information Center, Sydney, Australia

Point mutations in the SOX9 HMG protein lead to a skeletal malformation syndrome called *campomelic dysplasia*. With reference to a model of the SOX9 HMG protein domain, Dr. Ranganathan discussed how point mutations in the protein affect DNA binding. In addition, the AGIC group has carried out a molecular modelling analysis of the two functionally-uncharacterized, two-domain whey acidic proteins.

Database Processing Using Multiple PCs Running Linux Supercomputing for All

Neil R. Taylor, BASF Bioresearch, Worcester, MA, USA

For scientists working with large amounts of chemical and biological data, both the quality and quantity of the research is closely related to computational throughput. Given the price of commercial high-end workstations, the construction of a high-end server consisting of multiple Intel-based PCs, running the freely available Unix-like operating system *Linux*, offers a price/performance ratio far superior to current standards in the biotechnology and pharmaceutical industries. Dr. Taylor discussed how BASF assembled their PC cluster from off-the-shelf components and overcame the problems of parallel computing using the Python scripting language and the interface program PVM. He then demonstrated how they could use this resource to perform high-end 3D structure-based searching of several databases quickly with a view to structure-based drug discovery.

A Method for Computational Combinatorial Peptide Design of Inhibitors of the Ras Protein

Herbert R. Treutlein, Ludwig Institute for Cancer Research, CRC for Cellular Growth Factors, Parkville, Australia

A three-step computational approach, employing a "Multiple Copy Simultaneous Search" technique, was proposed for the design of a peptide inhibitor for the Ras protein. The advantages and limitations of the new technique were discussed.

Rapid Non-empirical Methods for Estimating Relative Binding Affinities

Alan E. Mark, Laboratory of Biophysical Chemistry, University of Groningen, The Netherlands

At present, the computational screening of potential ligands to fit drug targets relies on empirically-based discrimination functions. However, the free energy of a system is a global property and as such cannot be expressed simply as a sum of terms. According to Mark, free energy perturbation based approaches may, in principle, be used to estimate free energy changes but they are too computationally expensive for most applications, as a new simulation has to be initiated for each new compound of interest. The Mark group has, however, described a series of methods that can estimate the binding free energy for multiple compounds from a single simulation. Using appropriately biased "reference ensembles," for example, the binding free energies of multiple compounds could be computed from a pre-stored trajectory. This new approach should allow the rapid screening of newly proposed compounds for drug/protein binding free energies with an increase in efficiency of several orders of magnitude over traditional free energy methods.

The Design of Protein-interaction Focused and Conformationally Constrained Libraries

Tran Trung Tran, Center for Drug Design and Development, University of Queensland, Queensland, Australia

Protein-protein interactions remain one of the most difficult molecular recognition events to model. The affinity and specificity of these interactions exist as a function of both the amino acid composition of the interacting surfaces (chemical diversity) and the 3-dimensional arrangement of the residues on the interacting surface of the protein (conformational diversity). Previously, the Brisbane group have reported strategies for exploring and mimicking the conformational space of proteins, by classifying and matching positions and orientations of the C α -C β bonds in the amino acid side chains residing on the surface. In this presentation they reported the application of this strategy to β -turns with the development of a program to screen structural databases for relevant scaffolds.

The Fine Lines Between Interpretability, Relevance, and Predictive Utility of Variables in QSAR Studies

Chris L. Walker, Sphinx Pharmaceuticals, NC 27709 USA

Dr. Walker discussed how QSAR modelling is plagued with uncertainties. The increase in the number of molecular

descriptors (firm, intermediate or soft) has given rise to numerous QSAR algorithms, which vary with respect to variable selection. In his lecture he discussed the questions of variable selection and descriptor choice within the context of a study designed to identify potential environmental estrogens. He then extended his discussion into the area of molecular design and addressed the limits of interpolation and extrapolation of QSAR models.

Prediction of Molecular Properties from Momentum-Space Wavefunctions

Matthew Sykes, Flinders University, Adelaide, Australia

The Flinders group has calculated various parameters that characterize the momentum-space electron density distributions for a wide range of small molecules, and have examined their relationship to molecular properties. They reported success in the prediction of polarizabilities, hyperpolarizabilities, and critical molecular volumes when using parameters that are dependent upon density distribution near the momentum-space origin. When applying these techniques to the analysis of anaesthetic potency data they have obtained a high level of agreement between prediction and experiment.

The Development of Models for Clozapine Bound to Dopamine Receptors

D.K. Chalmers, Victorian College of Pharmacy, Monash University, Parkville, Australia

The antipsychotic clozapine is used widely to treat drug resistant schizophrenia and is thought to act by agonism of the dopamine D₄ and serotonin 5HT₂ receptors. The Chalmers group has developed drug receptor models for the binding of clozapine, and other ligands, to the transmembrane regions of the five-dopamine receptor subtypes (D₁-D₅). They have constructed drug-receptor models based on the older rhodopsin model of Herzyk and Hubbard using the program MODELLER and are using structure activity relationship data to position the ligand. It is anticipated that this approach will be used for the design of new dopamine ligands.

Docking-Derived Pharmacophores

Renate Griffith, University of Wollongong, Wollongong, Australia

The Wollongong group has constructed models of the α_{1A} and α_{1B} subtypes of the adrenergic receptor subfamily, an important family of membrane coupled receptors for which there is currently a paucity of structural data available. They have docked the natural ligand adrenaline and also a model of a rigid α_{1A} selective agonist into these models, and have thus been able to propose a new interaction between adrenaline and the α_{1A} protein. They have constructed so-called "docking derived pharmacophores" from the ligand receptor complex models. This new approach offers further opportunities for ligand design.

Design and Modelling of Sulfoxide and Sulfinamide Peptidyl-Prolyl Isomerase Inhibitors

Peter Karuso, Macquarie University, Sydney, Australia

Using both modelled and solution state structures of the FK506 immunosuppressor protein and its receptor FKBP, the Karuso group has targeted the key non-bonding interaction between a carbonyl group and several aromatic hydrogens in the natural guest in their design of sulfur-based isosteres. *Ab initio* computational techniques (at the HF/6-31G*/MP2/tzvp level of theory) were used to assess the strength of potential non-covalent interaction occurring between the sulfoxide/sulfinamide function and the aromatic hydrogens. Testing of the derived materials suggests that even simple sulfoxides and sulfinamides are reasonable inhibitors of FKBP.

Towards a Fundamental Understanding of Complex Industrial Processes via Molecular Modelling

A.L. Rohl, Curtin University, Perth, Australia

The Rohl lecture discussed possible applications of advanced molecular modelling techniques in the minerals processing industry. Specifically, the Curtin group is currently attempting to gain an atomic level understanding of the Bayer process using both atom-atom potential and QM calculations. The effects of cations on the aluminium hydroxide precipitate were also discussed.

Towards First Principles Calculations on Complex Materials

Julian D. Gale, Imperial College, London, UK

In his lecture, Dr. Gale discussed how a new program — SIESTA — has been used to model thousands of atoms in complex materials using density functional methods both in the gas and condensed phases. The size of the system is made possible by designing the key operations such that the cost of the calculation can scale linearly with the system size. He then showed how this technique could be applied to both organic and inorganic systems.

Industrial Applications of Molecular Dynamics Simulation of Liquid-Crystalline Polymers

Pavel D. Ball, RMIT University, Melbourne, Australia

Using MSI's Cerius program, the RMIT group has demonstrated how molecular dynamics calculations can be used to model liquid crystalline polymers. The group has used the technique to predict structural properties, calculate the order parameter, analyze orientational properties as well as predict other key physical properties such as decomposition and glass transition temperatures.

First Principles Simulation of Galena (PbS)

Joseph Muscat, CSIRO Minerals, Australia

Dr. Muscat reported the preliminary results of a study of the industrially important mineral sulfide, Galena, using Hartree-Fock and density functional theory. The CSIRO group has investigated the structural and electronic properties of the bulk material. Its stable low index surface is, apparently, an essential prerequisite for subsequent modelling of the interaction of flotation collectors with sulfide surfaces.

COMP HISTORY

History of the Formation of the ACS Division of Computers in Chemistry

Peter Lykos

*Illinois Institute of Technology, Life Sciences Building,
Chicago, IL 60616. lykos@iit.edu*

This article summarizes the keynote address I gave, as the creator of COMP, at the Special Symposium Celebrating the 25th Anniversary of the Division of Computers in Chemistry held during the National Meeting of the American Chemical Society (ACS) in New Orleans, LA, in August 1999. I remain an interested observer in the evolution of COMP and do what I can to help promote its service to the chemical profession.

What follows is an attempt to review a few of the early computer-based enhancements to the practice of chemistry made visible via conferences and reports and that culminated in the creation of the ACS Division of Computers in Chemistry (COMP) in April 1974. Some follow-on activities spawned by the creation of COMP, revealing thereby some of its potential, are also described.

Before 3 April 1974

"Computers in Higher Education" was a report of the President's Science Advisory Committee that was published in February 1967. That Committee had almost 20 scientists on it and about three-fourths were chemists. They spawned a Panel on Computers in Higher Education that ginned up the referenced report. John Pierce of Bell Labs was the Chair. The report opened with the following:

"After growing wildly for years, the field of computing now appears to be approaching its infancy ... yet computers and computing have already fantastically increased our power to know as well as to do. They have made masses of

data that were previously intractable accessible to analysis and understanding. They have made it possible to trace the consequences of theories and assumptions in a wide diversity of fields."

The mid-1950s through the mid-1970s spanned a period of considerable growth and activity in several areas of computer-based enhancements to the practice of chemistry as well as impacts on a wide variety of constituencies. During that period, I was privileged to play several roles. This is an account of what happened during that period as that related to the formation of COMP in April 1974, as well as some direct consequences that followed shortly thereafter.

As a graduate student in chemistry at Carnegie Institute of Technology (that was soon to join with the Mellon Research Institute to form Carnegie Mellon University [CMU]), my research had to do with quantum chemistry and hence with solving secular equations. At that time, calculations were being done with mechanical table top calculators — read tedious. During the summer of 1953, I built an analog secular equation solver capable of handling the non-orthogonal problem for up to a six by six determinant.

But in 1954, IBM announced its IBM 650, the first mass-produced electronic digital computer that was affordable by universities. The IBM marketing staff predicted a market for 200 machines, but before the 650 ran its course, some 1800 were sold. Computational quantum chemistry for everyone had begun. By today's standards, the 650 was an extremely modest machine. The main memory was a rotating magnetic drum about 4 inches in diameter and about 12 inches long on the surface of which 2000 words could be stored. Input and output was via punched Hollerith cards. Data entry was done off-line with printing keypunches. Printing was done off-line with accounting machines such as the IBM 407 formattable by a wired plug board. Carnegie Mellon University became a focus for university use of the IBM 650 because of Alan Perlis. He was the designer of an algebraic compiler for the 650, The Internal Translator that was brought to CMU by the Graduate School of Industrial Administration so its students could have support for complex models of business and administration.

I ended up creating the academic program in computer science during the period 1964 to 1971, growing to 35 undergraduate and graduate courses with the fifth largest course enrollments in a 25 department university. In addition, we created the academic computer center that grew from an IBM 650 to an IBM 1620 to an IBM 7040/1401 to an IBM 7040/1301 (wherein was created SPOOLing) to an IBM 360/40 to a Univac 1108 (instead of an IBM 360/65). As the latter was not my choice, I left computing at IIT (still a professor of chemistry) and took a two-year leave of absence as a National Science Foundation (NSF) Rotator to create a new section, Computer Impact on Society.

During the computer-based activity at IIT, we became active in bringing the computer into academic programs. In the fall of 1959, I required our students taking Physical Chemistry lab to measure the vapor pressure of a pure liquid as a function of temperature and to punch their data into paper tape (flexowriter for the UNIVAC 1105 at the IIT

Research Institute). I then devised an experiment, Digital Computer Programming, where the students, in groups of 20, learned the concepts of addressable memory, 10 machine language instructions (in octal), and worked through an example that included using a logarithm subroutine (with operator controlled load and dump routines for input and output). Our labs went for three hours in the morning and three in the afternoon. In the afternoon class, we wrote a program to do a least squares fit of their data to a straight line. Many years later, the Charles Babbage Institute (CBI) called to inquire as to just when we first did that experiment and our records showed that first happened in November 1959. As well as the CBI folks could determine, that was the first instance where students learned to program a computer in a required course in a major discipline.

I invited a group of high school students to come to IIT for a few Saturdays. The high school students were highly motivated and well able to master programming in octal in machine language. So I expanded the program by inviting the superintendents at the over 100 high schools in metropolitan Chicago to nominate their very best for a Saturday morning program. That program was an instant success and over ten years some 15,000 high school students came to our campus to learn the rudiments of computer programming. As an aside, two students (a girl and a boy) from Roosevelt High School, were working on IIT's IBM 1401 and improved an IBM-supplied "multiply" subroutine as part of their first exposure to a digital computer!

Of course, their math teachers were not computer knowledgeable and came to me for a program for them. So we quickly created a Saturday program for high school teachers. Over ten years, some 1,200 teachers came to our campus to learn programming and went on to study computer applications. Out of that, we created the first ever Master of Science in Teaching – Computer Science (MST-CS). We acquired an IBM 360/40 so we could engender a remote job entry system so high schools could rent teletypewriters from the phone company and send batch programs to IIT over the phone lines using IITRAN, a student-oriented, extremely fast, and with comprehensive diagnostics, compiler. As we had created IITRAN it was duck soup to render all the key words and diagnostic messages in Spanish (for SPANTRAN) as well as in French, German, and Italian.

In parallel with all that high school activity, we created a one-day Saturday Conference on the Computer in Undergraduate Curricula, 25 January 1969, the first such national activity. Some 350 people came from 24 states, five people from the University of Iowa. Thus it was likely not a coincidence that the NSF supported a consortium based at the University of Iowa, some 17 months later, and conducted the second such conference (of what became a series of annual multi-day conferences with published and widely distributed proceedings focused on computers being applied in undergraduate curricula).

We moved on and became one of ten NSF supported programs across the U.S. having moved beyond providing computer support and computer programming skills. We took part in the most comprehensive and heavily funded Cooperative Venture in College Curriculum Development with

nine other colleges and universities in 1968-1971. Faculty were involved from Business and Economics, Education and Psychology, Sociology, Biology, Chemistry, Mathematics, and Physics. Indeed, *Chemical and Engineering News* magazine noticed what we were doing in Chemistry as part of that cooperative curriculum development project and wrote a cover story for the 4 November 1968 issue.

During the late 1960s, I was invited to become a member at large of the National Research Council and to chair a new standing committee in the National Academy of Sciences/National Research Council (NRC/NAS); namely, the NAS/NRC Committee on Computers in Chemistry. I presume that was the doing of Harrison Shull (Indiana University). It was an interesting confluence as my two-year association with the NSF overlapped with that membership and it was a short walk from the NSF to the NAS. Thus began a collegial and fruitful relationship with Martin Paul, Secretary of the Chemistry Division of the NAS. We proved to be an effective team.

Just before moving to Washington, DC, we designed the first week-long conference on Computers in Chemical Research and Education. That became the first of a biennial series that continues today. The program was a defining moment in the structure of computers in chemistry that constituted a blueprint and momentum for creating a new ACS Division. The week-long meeting took place at Northern Illinois University's Conference Center in July 1971. The following sessions were held:

- Data Acquisition and Refinement, Klaus Biemann, MIT
- Computer Assignment of Complicated Spectra, William Carnall, ANL
- Structure Determination and Chemical Kinetics, Walter Hamilton, Brookhaven National Laboratory
- Computer Assisted Instruction, Joseph Lagowski, University of Texas
- Theoretical Chemistry, Larry Snyder, Bell Labs
- Chemistry Curriculum Development, Joseph Denk
- Communications, Networks and Program Exchanges, David O. Harris, University of California, Santa Barbara
- Computer Control of Experiments, Roger Anderson, Lawrence Livermore National Laboratory
- Modern Methods of Handling Chemical Information, Martha Williams, IIT Research Institute

In addition, there were several "Birds of a Feather" sessions in the evenings including one on graphics.

Two weeks before the conference was to begin, an old friend from Yugoslavia dropped in and lamented the fact that his visa was about to expire so he could not attend. So I suggested to him we have a sequel in two years in Yugoslavia. And we did. Thus was started the biennial series. Subsequent International Conferences on Computers in Chemical Research and Education (ICCCRE) were conducted all over the world. The twelfth ICCCER took place at the University of Pune, India, in January of 1998 while the XIIIth is tentatively scheduled to be held in Berlin.

While with the NSF in Washington, DC, events transpired that added momentum to our computers and chem-

istry initiatives. Although I had been brought to the computer science division of the NSF to create a new section, I was temporarily assigned to the Computers in Research section where I soon discovered there were substantial monies allocated but a dearth of focused initiatives to which those might be applied. I began by posting little signs all over the office, "Think Hierarchical" and began encouraging proposals to interface lab computers and small workstations to departmental or campus-wide computer systems that in turn were linked to remote off site specialized or supercomputers.

That led me to visit Larry Roberts of the U.S. Department of Defense Advanced Research Projects Agency (ARPA) just a short cab ride away to learn more about the emerging ARPANET generally and how that technology might be brought to the service of computer augmented scientific research specifically. I did manage to get extremely favorable reviews for a project to involve developing hierarchical computing and had the cooperation of ARPA — including permission to acquire an Bolt Beranek Newman (BBN) IMP (interface message processor) needed to create an ARPANET node.

Unfortunately, interagency rivalry led to a special trip by a higher ranking NSF person to that site to advise the would-be PI to remove the IMP from his budget. Thus the NSF slowed down the process to the point where several years elapsed before the national network concept matured to an NSF initiative that could be (necessarily!) claimed as distinct from the ARPANET. As recently as 1995, over 20 years later, a Supercomputing Conference in San Diego staged a short-lived collaboration. Using novel communications protocols over high speed links, a small number of U.S. supercomputer centers performed computations cooperatively, passing data back and forth to create a giant meta-computer. However, 25 years after that first initiative, we still do not have an operational network linking universities for really large-scale scientific computing.

On the other hand, IBM has just announced a five-year, \$100 million R&D program to develop a PetaFLOPS (10^{15} FLoating point Operations Per Second) highly concurrent multiprocessor supercomputer, driven, ostensibly, by the need to help elucidate the mechanism of protein folding.

While at the NSF we continued an earlier initiative toward a National Center for Computation in Chemistry (NRCC) based on a preliminary design by Joe Hirschfelder (University of Wisconsin), Clemens Roothaan (University of Chicago), and myself. Working through the NAS/NRC Committee on Computers and Chemistry we were able to foster major conferences, national in scope and with published proceedings, where prominent researchers in several areas in chemistry described both cutting edge research and examples where progress was limited by lack of access to state of the art supercomputers. The first conference addressed theoretical chemistry but was followed by others such as computers and crystallography. Eventually the concept of a national center for computation in chemistry gathered momentum and with the interest of the U.S. Department of Energy (DOE) was moving slowly to a national

resource (not center) for computers in chemistry.

At that point, the Assistant Director for Physical Sciences of the NSF decided that the NSF should be involved. The DOE lab at Berkeley made a major and successful effort to win the competition, but it was clear their primary motivation was to shore up their CDC 7600 installation that was suffering from lack of support. That situation was made even more complicated because William Miller and Henry Schaefer III of the Chemistry Department, University of California, Berkeley, had pioneered in securing NSF support to purchase a minicomputer dedicated to theoretical chemistry research by way of a demonstration that would be more cost effective for that purpose than the CDC 7600.

The Miller/Schaefer proposal arrived in the NSF Chemistry Division where not much could be done as the money was not there but, more importantly, the change in policy for supporting computation for theoretical chemistry was not there. First, I needed to convince a very influential member of the Division of Computer Science Advisory Committee that supporting two theoretical chemists to acquire and to operate their own super minicomputer could prove to be a cost effective alternative to paying for computer time on a supercomputer of the day. We succeeded. I provided both the money and the guidelines whereby detailed records regarding cost/effectiveness had to be kept that would stand up under close scrutiny. University computing center directors were not going to take this challenge lying down! So Miller and Schaefer got their own super minicomputer and, as they say, the rest is history. A large number of chemistry departments acquired their own super minis and changed forever the mode of computer support in academic chemistry departments. That happened in industry as well.

During my transition from a severe case of "Potomac Fever" to life back in Chicago, Martin Paul and I put our heads together and decided to initiate a new Division in the American Chemical Society. We took advantage of the extensive mailing list of chemists in the National Academy of Sciences Chemistry Division and requested support of a petition to the ACS for the creation of a new ACS Division, the Division of Computers and Chemistry.

The responses started pouring in until we had collected almost 700 signatures of ACS members. Subsequently we distributed a profile and guidance questionnaire to the original petition signers and got an amazing response of 389. The range of chemists by sub-discipline was broad but was represented with significant concentrations of members of the following existing ACS Divisions: Organic, Inorganic, Analytical, Physical, and Education. Almost all the states of the U.S. were represented. The ACS required that, for it to be considered, a petition for a new Division needed to have 50 signatures. Our petition not only exceeded that threshold, but it had the largest number of petition signers for a new Division in the history of the ACS. When the recommendation was brought to the floor for vote by the ACS Council on 3 April 1974, there was little discussion. The vote to create the new Division was unanimously in favor.

Martin Paul and I selected a seven-person slate of offi-

cers and directors from the list of petition signers. We presented that recommendation to the ACS president who then formally appointed that slate. Given the short lead-time, we proceeded to create symposia and to recruit chairs for each for the next two national ACS meetings as follows:

September 1974

Larry Snyder (Bell Labs) organized a full day session with eight speakers commenting: "The National Academy of Science Report of a Feasibility Study of a National Center for Computation in Chemistry."

Bruce Kowalski (University of Washington) organized a three half-day symposium with the overall theme being "The Computer in Experimental Chemistry."

April 1975

Frank Harris (University of Utah) announced that he was calling for recommendations for a symposium on "Computer Program Certification and Transferability."

Edgar Meyer (Texas A&M) announced a call for papers for a symposium on "Computer Graphics, Input and Display."

During the process, the ACS notified us that our four-letter acronym would be CCHE to which we responded we would prefer COMP — and that was done. Thus not only was COMP born but it had hit the ground running. Thus began the process of learning how to operate a Division within the ACS structure.

After April 1974

We generated the first COMP Newsletter that was sent to every COMP petition signer on 5 August 1974. Mindful that that issue would serve an archival purpose, we summarized therein all the information relevant to the creation of COMP. Included therein was the original statement of purpose:

"Through symposia, programs and other means provide a forum for academia, government and industry to discuss and extend the large and growing impact of computer technology in all of chemistry; constitutes an interface between chemistry and computer science, mathematics, statistics, and engineering; examines computer, communication, servo/sensor and display hardware-software systems in areas such as real-time data acquisition, and control, chemical system modeling and simulation of bulk matter from a molecular perspective including quantum chemistry, computer-aided visualization and correlation of chemical structures with synthesis design, biological activity, etc., chemical data analysis including pattern recognition, information storage and retrieval, teaching chemistry and chemical engineering and managing all of chemistry-based endeavors."

Mindful of the need to develop focuses as snapshots in time of important and emerging areas of computers in chemistry, we organized a series of symposia conducted at national ACS meetings whose proceedings were published and thus distributed widely. The purpose was two-fold: to make available in printed form specific examples of how leaders in those specific areas were actually using computers to enhance the practice of chemistry and to encourage others

to develop corresponding COMP symposia with published proceedings.

That pulse of activity stimulated many others to work within COMP to organize and publish proceedings of computer in chemistry focuses.

By way of an ACS local section initiative, in March 1980, the Chicago Section of the ACS sponsored a conference "Use of Small Computers in the Chemistry Laboratory." The six paper proceedings were published (The Laboratory Computer Coursebook). A copy was sent to every one of the 5,000 members of the Chicago Section.

Meanwhile the Joint Board – Council Committee on Science of the ACS organized an intensive 24-hour (September 1983) total immersion conference for all the ACS division program directors by way of getting them to look more deeply into the future than what was being done at that time in planning technical programs for ACS National Meetings. We were invited speakers who were expected to deliver provocative lectures on emerging areas in the practice of chemistry; namely:

- George Pimentel, An Overview of Opportunities in Chemistry
- Monte Throdahl, Insights on Future Evolution of a Major Chemical Company (Monsanto)
- George Hammond, Insights on Future of a Major Company Whose Interests are Evolving in New Directions (Allied Corp.)
- Charles Kline, Breaking Through to New Technologies
- Peter Lykos, Insights into the Future and Impact of Computer Technologies.

That was a major step toward raising awareness within the ACS regarding the already large and rapidly-growing impact of the computer on chemistry.

One direct consequence of all the computers in chemistry activity — and the well published accounts thereof — came an invitation from the Division of Scientific Research and Higher Education, UNESCO, Paris, to do a study and analysis and a corresponding report, "The Computer's Role in Undergraduate Chemistry Curricula." By May 1977, 300 copies of the report had been sent to universities in developing countries through its regional offices in Africa, South Central Asia, Southeast Asia, the Arab States, and Latin America. In addition, 500 copies were sent in response to individual requests — mostly from the United States, Canada, and Europe. Thus our international outreach via the ICC-CRE was augmented in an interesting manner.

Book REVIEW

Advances in Quantitative Structure-Property Relationships, Volume 2

Edited by Marvin Charton (Pratt Institute, Brooklyn, NY) and Barbara I. Charton (St. John's University Science Library). JAI Press, Stamford, CT, 1999. ix + 257 pp. Hardcover \$109.50. ISBN 0-7623-0067-1.

This is a book the QSAR purists will love. The book focuses on the use of traditional QSAR techniques such as regression analysis to elucidate research problems on a far-ranging scientific landscape.

Quantitative structure-property relationships (QSPR) and quantitative structure-activity relationships (QSAR) have been discussed at the biennial Gordon Research Conferences on QSAR since the conferences started in 1975. These are well attended and popular gatherings. Since the early 1980s, however, a common complaint expressed by some of the "elderly bishops" of QSAR has been that the conferences included too much molecular modeling, whereas pure QSAR topics like regression analysis got short shrift. On the other hand, the "young Turks" are using molecular modeling, QSAR/QSPR, structure-based design, docking, artificial intelligence, or whatever other computational chemistry tools are useful for answering the research questions at hand. Regardless of the complaints, it should be noted that the majority of the scientists involved in the QSAR Gordon Research Conferences have been magnanimous toward the infusion of modeling because of its utility in understanding relationships between chemical structures and observed bioactivities or other properties.

The QSAR purists will be glad to know that this book is free of modeling influences. The authors of the four chapters provide plenty of regression equations and lots of tables of data. Generally, the level of presentation is appropriate for the expert. It was apparently assumed by the authors that the reader would have a good understanding of physical organic chemistry and even prior knowledge of an occasional descriptor or two.

Little is said in the book about the basic motivation of the QSAR/QSPR approach, i.e., the scenario where the researcher is presented with a table of chemical structures and properties and asked to explain it. Often, little or nothing is known about the association and transport processes that impinge on the observed properties. It is the job of the QSAR expert to analyze the data to fathom what processes may be involved. To do a QSAR analysis, a pretty standard approach is employed. The set of molecules in the study is described in terms of descriptors measuring the steric, electronic, and lipophilic characteristics of the whole molecules or substituents therein. Then statistical regression analyses are done to discern relationships between the properties and descriptors. If statistically strong relationships are found, then voilà!, the computations may be publishable. A fine art has developed over the years on how to interpret the equations in chemically useful ways.

Chapter 1, entitled "Exploring the Energetics of Binding in Chromatography and Related Events," is by Philip S. Magee, who for the last 16 years has been a consultant in QSAR after a long career at Chevron Chemical's Ag Division (Ortho) in California. In 33 pages, he analyzes adsorption energies and retention times for sets of compounds on a variety of substrates. The latter include inorganic materials (clays, silica, and alumina), organic materials (cellulose and activated carbon), and mixtures (soil). Among the compounds whose separation is discussed are aromatic hydrocarbons, nitrobenzenes, alcohols, carboxylic acids,

phenols, pesticides, and lots of pyridines. Many of the findings presented are the author's own previously unpublished work. With each regression equation presented, the author thoughtfully interprets the meaning of the descriptors that are found useful. In so doing, he concludes that various types of intermolecular forces come into play in the adsorption process depending on the compound series and substrate. These forces manifest themselves through a diverse set of dispersion, steric, hydrogen-bonding, and electronic descriptors. The most frequently cited publications are the *Journal of the Chemical Society, Faraday Transactions*, and the *Journal of Physical Chemistry*. The distribution of the references by year is shown in Figure 1.

The second chapter is on "Structural Effects on Gas-Phase Reactivities." The authors are collaborators from Spain, Venezuela, and Japan: Gabriel Chuchani, Masaaki Mishima, Rafael Notario, and José-Luis M. Abboud. The authors examine reactions involving reagents and products that are ionic or neutral. Starting with the original sigma of Hammett's famous equation relating acidities of substituted benzoic acids to the electronic effects of the substituents, the sigmas have blossomed to σ_D , σ_I , σ_R , σ_R^+ , σ_R^- , σ° , σ^+ , σ^- , σ , and so on. The authors analyze a full panoply of reactions for which the electronic effects of the substituents are influential on rate or equilibrium. This longest chapter of the volume (92 pages) could have benefited by having a section of conclusions and a summary of lessons learned from the regression analyses. Many of the authors' own papers are cited because they have contributed much to this area of research. The most frequently cited publications are the *Journal of the American Chemical Society* and the *Journal of the Chemical Society, Perkin Transactions 2*. The distribution of references by year is shown in Figure 2.

Chapter 3, entitled "The Prediction of Melting Point," is a tour de force by John C. Deardon. In an engaging style, he explains why it is of interest to be able to predict melting points: drug action and toxicity depend on solubility; solubility is related to melting point and lipophilicity. The scientific study of melting points has a long history. He traces the literature all the way back to the 1870s. Over the years, various investigators have used descriptors such as atom counts, density, vibrational frequency, molecular weight, conformational flexibility, and additive constitutive group parameters. More than 80 empirical relationships and regression equations are given for predicting melting points. Most studies have been on hydrocarbons, but some more interesting molecules are also reviewed. Areas of research still needing attention are pointed out. Although the problem of fitting the melting points of homologous hydrocarbons to various equations appears well under control, the treatment of drug-like molecules remains a challenge. For the latter type of compounds, the average error in the predictions amounts to over 25 degrees. There are surprisingly few computer programs available for predicting melting points; limitations of these programs are pointed out. The most frequently cited publication is the *Journal of the American Chemical Society*. The distribution of references by year (Figure 3) is clearly different from that of the other chapters because of the area's long history.

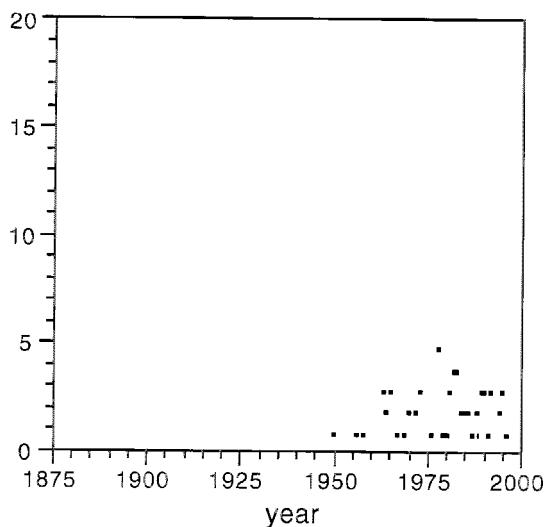


Figure 1. Publication year for references cited in Chapter 1.

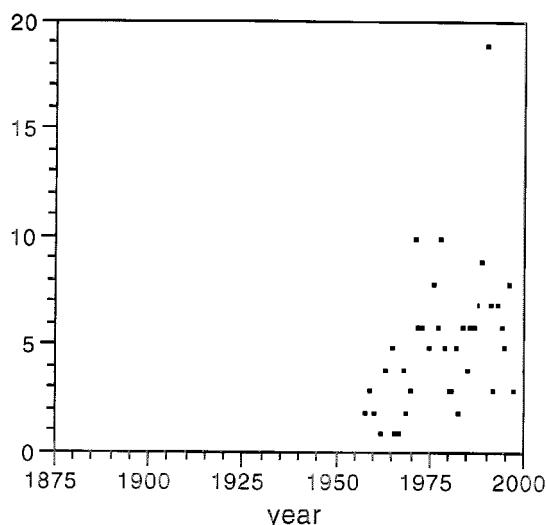


Figure 2. Publication year for references cited in Chapter 2.

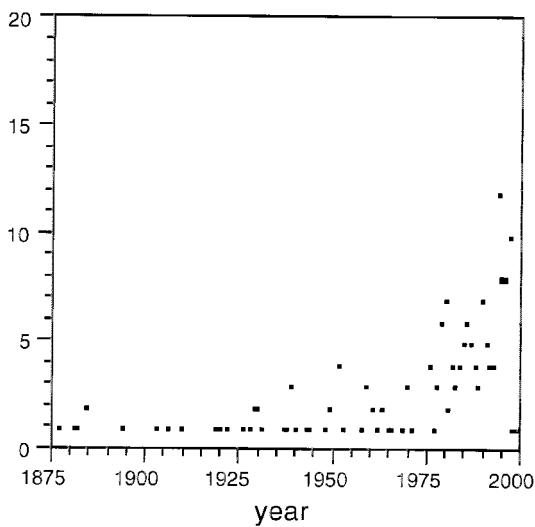


Figure 3. Publication year for references cited in Chapter 3.

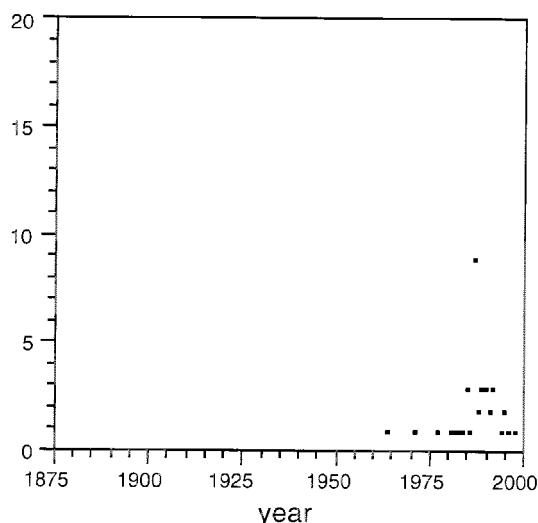


Figure 4. Publication year for references cited in Chapter 4.

Chapters 4 is on “The Application of the Intermolecular Force Model to Peptide and Protein QSAR.” The author is one of the co-editors, Marvin Charton. He relates the presence or absence of substructures to binding constants, IC_{50} values, and k_{cat} values. Among the molecular systems treated are analogs of peptide renin inhibitors, angiotensinogen, pepstatin, human growth hormone, subtilisin, hirudin, thymidylate synthase, and glutamyl-tRNA synthase. The reader will have to follow closely the large number of abbreviations introduced. At the end of this chapter is a glossary of abbreviations, but the reader may wonder why they are not in alphabetical order or why some abbreviations were not included. The most frequently cited publication is the

Journal of Medicinal Chemistry. The distribution of references by year is shown in Figure 4.

The book has an excellent appearance; production standards were high. One wonders, however, why all the regression equations and associated statistics are not presented in a uniform style from chapter to chapter or why the word collinear is not always spelled with two l's. Curiously, the Wiley-VCH journal, *Quantitative Structure-Activity Relationships*, which is affiliated with the QSAR and Modelling Society, is cited only about ten times in the entire book. The low overlap of work described in this book and work published in the journal is odd. Besides what the low overlap says about the book and the journal, it also means that

every QSAR researcher will need both in order to have on hand a full complement of literature. No QSAR library would be complete without this book. Priced at about US\$ 0.42 per page, the book is targeted at libraries and specialist individual buyers.

The editors are to be commended for a superb selection of chapters. This book should be well received by the QSAR community. The book will advance the field of computational chemistry by showing how calculations by traditional QSAR approaches are shedding light on diverse chemical questions.

*Donald B. Boyd, Ph.D., Department of Chemistry, Indiana University-Purdue University at Indianapolis (IUPUI), 402 North Blackford St., Indianapolis, IN 46202-3274.
boyd@chem.iupui.edu*

MEETING ABSTRACTS

Abstracts for the 3rd Young Modellers Forum, Nov 26th 1999

Construction of a Full 3D Model of the Transpeptidase Domain of *S. pneumoniae* PBP2x

P.A.M. Van Hooft and H.-D. Höltje, Institut für Pharmazeutische Chemie, Heinrich-Heine Universität Düsseldorf, Universitätsstrasse 1, D-40225 Düsseldorf, Germany

To date, *S. pneumoniae* PBP2x is the only high molecular mass penicillin-binding protein that has been crystallized.¹ Due to the low resolution (3.5 Å) of the solved X-ray structure,² however, only the coordinates of the Cα-atoms have been deposited in the Brookhaven Protein Data Bank (1PMD). Here, a new approach making use of local alignments and geometrical modifications will be presented, in order to generate full 3D coordinates of the transpeptidase domain. In addition, it will be shown how the model, including a structural water molecule,³ was relaxed with simulated annealing and molecular dynamics techniques using InsightII/Discover.

Meanwhile, the structure has been solved to 2.4 Å resolution. Therefore, the model was sent to Dr. Dideberg for verification. Apart from a loop region next to the active site (T370-M386), the model appeared to be correct. We obtained new X-ray coordinates for this region.⁴ After incorporating these new coordinates into the model, the relaxation procedure was repeated. Finally, benzylpenicillin was docked into the active site. The stability of the complex will be shown by a molecular dynamic simulation.

References

1. P. Charlier et al., J. Mol. Biol. 232 (1993) 1007.
2. S. Pares et al., Nature Struct. Biol. 3 (1996) 284.
3. N. Mouz et al., Proc. Natl. Acad. Sci. USA 95 (1998) 13403.
4. O. Dideberg, personal communication.

Digitally Filtered Molecular Dynamics: The Frequency Specific Control of Molecular Dynamics Simulations

S.C. Phillips, C.M. Edge and J.W. Essex, Department of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

A new method for modifying the course of a molecular dynamics computer simulation is presented, based on the selectively enhanced molecular dynamics (SEMD) method of Dauber-Osguthorpe et al.¹ Digitally Filtered Molecular Dynamics (DFMD) applies the well established theory of digital filters to molecular dynamics simulations, enabling atomic motion to be enhanced or suppressed in a selective manner solely on the basis of frequency. The basic theory of digital filters and its application to molecular dynamics simulations is presented, together with the application of DFMD to the simple systems of single molecules of water and butane. The high degree of selectivity and control offered by DFMD, and its ability to enhance the rate of conformational change in butane, is demonstrated. The extensions required to apply this methodology to enhancing conformational change in biological molecules in the condensed phase will be described.

Reference

1. P. Dauber-Osguthorpe, C.M. Maunder, and D.J. Osguthorpe. J. 1996. Comput. Aided Mol. Design 10:177.

Development of Functional Threading Templates for Improved Protein Fold Recognition

C. Hadley and D.T. Jones, University of Warwick, Warwick UK

Threading is based on the observation that proteins with no discernible sequence similarity may still share the same fold. A sequence of unknown structure ("target") is threaded onto a library of known folds ("templates") and the energy (or "goodness") of each fit is calculated. The template offering the lowest energy model is assumed to have the best matching fold. THREADER is one of the most successful threading algorithms developed to date (Jones et al., 1992).

In attempts to improve the accuracy of THREADER, we focus on the fold template library, against which sequences are threaded. The current library is derived from the CATH structure classification database (Orengo et al., 1997). To evaluate the reliability and coverage of the fold library, a systematic comparison of three protein structure classification databases (SCOP, CATH and FSSP) has been completed (Hadley and Jones, 1999). This analysis has pinpointed various issues that need to be addressed when compiling future template libraries from the available fold classification resources.

Threading success may be further improved by integrating functional information into the templates. Current work involves combining knowledge-based information (known functional sites) with empirical information (func-

tion prediction using residue conservation and structural information). The improvement in THREADER accuracy as a result of template development is currently being assessed.

References

- Hadley, C. and D.T. Jones. 1999. A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. *Structure* 7:1099-1112.
- Jones, D.T., W.R. Taylor, and J.M. Thornton. 1992. A new approach to protein fold recognition. *Nature* 358:86-89.
- Orengo, C.A., A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, and J.M. Thornton. 1997. CATH – a hierachic classification of protein domain structures. *Structure* 5:1093-1108.

An Approach to Improve Multiple Alignments of Protein Sequences Using Predicted Secondary Structure

Andy Jennings, Discovery Chemistry, SmithKline Beecham Pharmaceuticals, New Frontiers Science Park, Third Avenue, Harlow, Essex CM19 5AW, UK

The object of this work was to improve multiple sequence alignments. A method is described where the secondary structure of proteins is predicted and this information, coupled with a simplified description of the amino acids, is used to produce multiple sequence alignments. This method improved the accuracy of the resulting alignments by between 5 and 14 percent when compared to full sequence profile alignments (as scored against structural alignments). These improved alignments were used to predict the secondary structure of the sequences they contain. The predictions based upon these more accurate alignments were more accurate than those produced from less optimal alignments and showed an improvement of 6 percent for a 3-state (helix, sheet and coil) prediction. The method uses entirely public domain software and can therefore be implemented by any scientist in the field.

Intermolecular Interactions on Ligand Binding

James A. Lumley, Department of Chemistry, University of Reading, Reading, UK. J.A.lumley@reading.ac.uk

The Multiple Copy Simultaneous Search (MCSS) Technique¹ has been designed to calculate functionality maps of binding sites by simultaneous minimization of several thousand functional group minima in a localized region of a target structure. We report here the use of this technique to probe the α -subunit binding site of the tryptophan synthase multienzyme complex. Several crystal structures are available which represent the enzyme in its native and ligand bound conformations. A D-h-2 hinge closure mechanism is seen on ligand binding with closure of Loop-6 to isolate the active site and promote interactions between several key catalytic residues. MCSS results are obtained for the differing structures to identify the importance of possible hydrophobic collapse of the binding site around bound ligands and its implications in later structure based ratio-

nal design. MCSS results for a series of single and fused ring systems are compared against known IC50 values to assess the correlation between theoretical binding interaction energies and IC50 values. Conclusions are drawn that identify residues key to the efficient binding of ligands, and MD simulations are used to explore their role in the binding of various newly proposed inhibitors of the active site. The relative contribution of hydrophobic interactions and hydrogen bonds to ligand binding are considered. These insights are being used for the analysis of current synthetic targets and the proposal of further synthetically viable inhibitors.

Reference

- Miranker, A. and M. Karplus. 1991. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 11:29-34.

Modelling the Loading Rate Dependence of Ligand Interaction Forces

Philip M. Williams, Eoin Galligan, Sarah Harris, Adam Moore, Molly M. Stevens, Lisa H Pope, Stephanie Allen, Martyn C. Davies, Clive J. Roberts, and Saul J.B. Tendler, Laboratory of Biophysics and Surface Analysis, School of Pharmaceutical Sciences, The University of Nottingham, University Park, Nottingham NG7 2RD, UK

Nature has evolved pairs of molecular species that interact and withstand forces dynamically over a range of stress rates. This hyper-variability bestows properties on the strength of the interaction that permits structure to be formed depending on the mechanical environment of the complex. This dynamic strength of adhesive interactions has been probed using a variety of biophysical techniques including the biomembrane force probe (BFP) and the atomic force microscope (AFM). In order to investigate the role of structure in this mechanical behavior, it is necessary to study the molecular behaviour on the atomic scale over the range of timescale and stresses. Traditional methods of such analysis, using computational molecular dynamics procedures, suffer in that the time available for exploration is in the order of nanoseconds. Low rates of force loading, such as those probed naturally and experimentally, are therefore out of reach of the analysis. Here, we show how a combination of traditional computational methods and transition state simulation through Monte Carlo methods can extend the time over which a system can be analysed.

Prediction of Protein Functional Sites: The Evolutionary Trace Method

Mark K. Dean, Department of Biological Sciences, Central Campus, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK

The evolutionary trace method (Lichtarge et al., Proc. Natl. Acad. Sci. USA, 93, 7507, 1996) is a sequence based approach to predicting functional sites on proteins. We will outline developments of this powerful method and describe its application to SH2 domains, cytochrome c oxidase and

to G-protein coupled receptors (GPCRs). Control calculations on the SH2 domain identified the peptide binding site, as reported previously. For cytochrome c oxidase, the method identified the cytochrome c binding site on sub-unit II, as determined by site-directed mutagenesis experiments. For each GPCR family or subfamily, the method predicted the occurrence of functionally important clusters of residues on helices 5 and 6 and on helices 2 and 3. The cluster on helices 5 and 6 is consistent with previous knowledge of dimerization in GPCRs (Gouldson et al., Protein Eng., 11, 1181, 1999) and implies that all GPCRs have the potential to dimerize. The cluster on helices 2 and 3 (Gkoutos et al., Int. J. Quant. Chem. Biophys. Quarterly, 74, 371, 1999) is of unknown function and some possible explanations will be given.

Poisson-Boltzmann Studies on the Mechanism of Nitrite Reductase

Richard I. Maurer, Department of Biological Chemistry, Central Campus, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

Copper nitrite reductase is an important enzyme in the denitrification cycle. Its role is to catalyse the reduction of nitrite to nitric oxide through a one-electron reduction. Currently there is great interest in understanding the mechanism, partly with a view to designing biomimetic systems. Key steps in the mechanism involve protonation of the nitrite and reduction of the copper II site to which the nitrite is bound. However, there is considerable uncertainty as to the origin of the proton donor and this precludes a systematic hybrid QM/MM study — despite good structures obtained using MNDO for some copper nitrite model compounds. Consequently, we have used the Poisson-Boltzmann method to study the electrostatic properties of the enzyme and to determine the pK_a s. The pK_a s were found to be dependent on the protein microenvironment, including the oxidation state of the copper sites and the protein dielectric. The use of a protein dielectric constant equal to that of the solvent, rather than usual protein dielectric constant of about 2-4, has been shown to give better computed pK_a s for most titratable residues. This approach was used here and its justification is discussed. The proton donor in the *alcaligenes faecalis* nitrite reductase was found to be Asp 98 rather than His 255. The mechanism is discussed in the light of these new theoretical results.

Molecular Discrimination Between Homologous Binding Sites

James Smith and P.M. Dean, Drug Design Group, University of Cambridge, Department of Pharmacology, Tennis Court Road, Cambridge CB2 1QJ, UK. js252@cam.ac.uk

Structural diversity within functionally similar binding sites has received little attention and almost no systematic analysis in relation to the problem of drug design. Systematic comparisons of functionally similar binding sites can reveal the identification and characterization of discriminatory

interactions. Conventional approaches exploit homology modelling methodology namely 3D alignment and superposition of the C backbones. However, this is limited to shape similarity and not local functionality; problems arise when considering sites with no obvious correspondence between atoms, or when the ligands have different binding modes and conformations, perhaps during catalytic activity.

We have developed a new approach based on superposing ligand structures after they have been unambiguously divided into conformational classes. Forty-two representative purine diphosphates (ADP & GDP) and their mimics, co-crystallized with their binding domains from the Protein Data Bank, were divided into seven significantly different conformational groups. Seven representative ligand conformation were then used as reference frames for the superposition of the sites. The superposition of C atom positions produced not only the common architecture but also the similar and dissimilar contact residues.

Ligand Protein Contact data¹ was incorporated allowing the identification and scoring of putative hydrogen bonds for each ligand atom. Those that simultaneously formed hydrogen bonds to multiple receptor atoms were given greater weight, therefore ranking the selection of site points needed for any future pharmacophore design. Not only was it possible to prioritize the interactions that discriminated within the conformational groups, but also possible to discriminate between each of the binding site conformations. The C α atom superpositions in each conformational group also gave distinct alignments of discontiguous sequences that provided a clear foundation for further 3D functional motif searching.

This development provided rapid discrimination between functionally similar sets of binding sites and should improve the efficiency of directing de novo drug design.

Reference

- Sobolev, V., Wade, R., Vriend, G., and Edeman, M. (1996). Molecular Docking Using Surface Complementarity. PROTEINS: Structure, Function and Genetics, 25, 120-129.

QSAR in the Prediction of DNA Damaging Capacity of a Group of Congeneric 1,4-Naphthoquinones

Jane L. Roberts,¹ R.F. Bilton,² J.C. Dearden,^{1,1} School of Pharmacy and Chemistry, ²School of Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

Quantitative Structure Activity Relationship studies (QSARs) not only provide a prediction of biological activity for a series of compounds but they may also provide useful insights into the mechanisms or modes of action that influence their activities. Simply put QSAR analysis is an approach which statistically relates one or more physico-chemical or structural parameters of similar compounds to a quantitative measure of the biological activity of those compounds.¹⁻⁷

Quinones are ubiquitous in nature and have important biological functions including roles in oxidative phospho-

rylation and electron transfer.⁸⁻¹⁰ They are widely used for a variety of functions including anticancer, antibacterial, antifungal and antimalarial agents. A characteristic feature of the quinone moiety is its ability to undergo reversible oxidation-reduction and form semiquinone and oxygen radicals e.g. superoxide (O_2^-).^{10,11} It is the enzymatic reduction and oxidation of the quinone under aerobic conditions leading to the production of the semiquinone radical which is termed redox-cycling and is the reaction of interest in our study. It has been demonstrated that synthetic quinones, such as menadione, cause DNA damage in different cell systems, possibly being mediated by free radicals generated during redox cycling.^{12,13} The 1,4-naphthoquinones (1,4-NQs) redox-cycle to varying degrees^{14,15} and it is the purpose of this study to investigate the varying redox-cycling abilities of these variously substituted 1,4-NQs and thus their varying DNA-damaging capacities.

References

1. Benigni, R., Passerini, L., Gallo, G., Giorgi, F., and Cotta-Ramusino, M. 1998. Environ. Mol. Mutagen. 32:75-83.
2. Benigni, R. and Giuliani, A. 1994. Mutat. Res. 306:181-186.
3. Benigni, R., Andreoli, C., and Giuliani, A. 1989. Carcinogenesis 10:55-61.
4. Cronin, M.T.D. and Dearden, J.C. 1995. Quant. Struct.-Act Relat. 14:117-120.
5. Cronin, M.T.D. and Dearden, J.C. 1995. Quant. Struct.-Act Relat. 14:329-334.
6. Hansch, C., Telzer, B.R., and Zhang, L. 1995. Crit. Rev. Toxicol. 25:67-89.
7. Cronin, M.T.D. 1996. Toxicol. in Vitro, 10:103-110.
8. Monks, T.J., Hanzlik, R.P., Cohen, G.M., Ross, D., and Graham, D.G. 1992. Toxicol. Appl. Pharmacol. 112:2-16.
9. O'Brien, P.J. 1991. Chem. Biol. Interact. 80:1-41, published erratum appears in Chem. Biol. Interact. 1992. 8:219.
10. Powis, G. 1989. Free Radic. Biol. Med. 6:63-101.
11. Roginsky, V.A., Barsukova, T.K., and Stegmann, H.B. 1999. Chem. Biol. Interact. 121:177-197.
12. D'Odorico, A., Sturniolo, G.C., Bilton, R.F., Morris, A.I., Gilmore, I.T., and Naccarato, R. 1997. Carcinogenesis 18:43-46.
13. Woods, J.A., Young, A.J., Gilmore, I.T., Morris, A., and Bilton, R.F. 1997. Free Radic. Res. 26:113-124.
14. Munday, R., Smith, B.L., and Munday, C.M. 1999. Chem. Biol. Interact. 117:241-256.
15. Munday, R. 1997. Autoxidation of naphthohydroquinones: effects of metals, chelating agents, and superoxide dismutase. Free Radic. Biol. Med. 22:689-695.

Benchmarking PSI-BLAST in Genome Annotation

*Arne Müller, Robert M. MacCallum, and Michael J.E. Sternberg, Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.
a.mueller@icrf.icnet.uk*

The recognition of remote protein homologies is a major aspect of the structural and functional annotation of newly determined genomes. In this work (Müller et al., 1999) we benchmark the coverage and error rate in genome annotation using the state of the art homology-searching pro-

gram PSI-BLAST (position-specific iterated basic alignment tool) (Altschul et al., 1997). We evaluate the *one-to-many* success rate for homology based recognition, as often there are several homologues in the database and only one needs to be identified for annotating the sequence. In contrast, previous benchmarks considered *one-to-one* recognition in which a single query was required to find a particular target (Park et al., 1998). The benchmark constructs a model genome from the full sequences of the structural classification of protein (SCOP) database (Murzin et al., 1995) and searches against a target library of remote homologous domains (<20% identity) of known structure. The structural benchmark provides a reliable list of correct and false homology assignments. PSI-BLAST successfully annotated 40% of the domains in the model genome that had at least one homologue in the target library. This coverage is more than three times that if *one-to-one* recognition is evaluated (11% coverage of relationships between all possible relationships). Although a structural benchmark was used, the results equally apply to just sequence homology searches. Accordingly, structural and sequence assignments were made to the sequences of *Mycoplasma genitalium* and *Mycobacterium tuberculosis* (see <http://www.bmm.icnet.uk>). The extent of missed assignments and of new superfamilies can be estimated for these genomes for both structural and functional annotations.

References

1. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. Nucleic Acids Res. 25:3389-3402.
2. Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of protein databases for the investigation of sequences and structures. J. Mol. Biol. 247:536-540.
3. Müller, A., MacCallum, R. and Sternberg, M.J.E. 1999. J. Mol. Biol. (in press).
4. Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T., and Chothia, C. 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. J. Mol. Biol. 284:1201-1210.

A Linear Interaction Energy Study of a Series of Neuraminidase Inhibitors

Ian Wall, Southampton University, Highfield, Southampton, SO17 1BJ, UK

The Linear Interaction Energy (LIE) method, has been applied to the calculation of the binding free energies of 15 inhibitors of the enzyme neuraminidase. This is a particularly challenging system for this methodology since the protein conformation and the number of tightly bound water molecules in the active site are known to change for different inhibitors. It is not clear that the basic LIE method will calculate the contributions to the binding free energies arising from these effects correctly. To determine whether it is appropriate to include extra terms in the LIE equation, a detailed statistical analysis was undertaken. Factor analysis was used to determine the number of useful dimensions

contained within the data, and hence the maximum number of variables to be considered when specifying a model or equation. Biased fitting methods using orthogonalized components were then used to generate the most predictive model. The final model gave a q^2 of 0.74, and contained van der Waals and electrostatic energy terms. This result was obtained without recourse to prior knowledge and was based solely on the information content of the data.

HOBO: an Alignment-free 3D QSAR Method

*Matthias Hennemann, Torsten Schindler, and Tim Clark,
Computer-Chemie-Centrum, Friedrich-Alexander Universität
Erlangen-Nürnberg, Nägelsbachstraße 25, D-91052 Erlangen,
Germany*

Most QSAR-methods such as CoMFA are based on the absolute positions of points in space and therefore require an alignment of the drug molecules. This alignment is to some extent arbitrary and is often a compute-intensive non-automatic procedure. The alignment problem can be overcome when relative coordinates are used to describe the topology of drug molecules. This requires a definition of points, which are then related to each other geometrically, as in spatial autocorrelation or pharmacophore methods.

We now present a method which uses distance and angle information to describe the relative positions of pharmacophore groups. In order to obtain a generally applicable method, we have used common binding features like hydrogen-bond acceptors and -donors and aromatic rings to define the binding sites in drug molecules. The detailed electrostatic and polarisability information available from semi-empirical MO calculations is used to describe these pharmacophore groups (Figure 1).

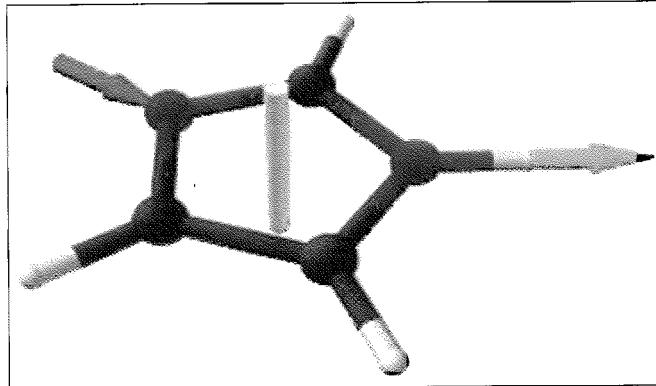


Figure 1.

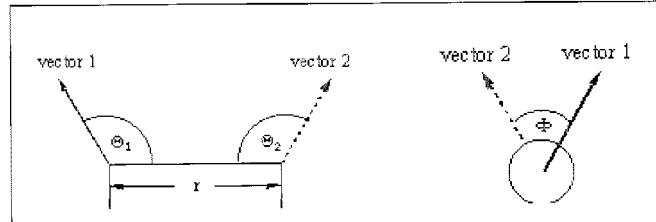


Figure 2.

The spatial relationship between the different sorts of binding sites on the molecule are defined by four parameters: the distance r , the two angles θ_1 and θ_2 between the vectors and the line connecting the two origins (Figure 2, left) and the dihedral angle Φ (Figure 2, right). This information was used to derive QSARs from limited (20-50 compound) activity data and 3D-pharmacophores from larger numbers of compounds that have been identified to be active. The pharmacophore models thus derived can be used to scan entire databases.

REVIEW ARTICLE

Rotational Superposition: A Review of Methods

Darren R. Flower

The Edward Jenner Institute for Vaccine Research, Compton, Newbury, Berkshire RG20 7NN, UK. darren.flower@jenner.ac.uk

Abstract

Rotational superposition is one of the most commonly used algorithms in molecular modelling. Many different methods of solving superposition have been suggested. Of these, methods based on the quaternion parameterization of rotation are fast, accurate, and robust. Quaternion parameterization-based methods cannot result in rotation inversion and do not have special cases such as co-linearity or coplanarity of points. Thus, quaternion parameterization-based methods are the best choice for rotational superposition applications.

Introduction

The alignment and superposition of molecules is a technique of fundamental utility in molecular modelling. Examples include the generation of pharmacophores and 3D database searching¹ or as a preliminary step to CoMFA.² Computational molecular biology has found uses for rotational superposition as diverse as backbone reconstruction, docking simulations, and in the systematic comparison of protein structures; whether used to effect database searching,³ to define structural superfamilies,^{4,5} or to draw inferences about the evolutionary relationships of proteins.⁶

Unsupervised molecular alignment determines which atoms are equivalent, while supervised molecular alignment optimally fits one vector set to another. Unsupervised methods are complex and highly heuristic in nature. Structural alignment methods, for large and small molecules, have proliferated. Brown et al. have published a review of methods for the unsupervised alignment of protein structures.⁷ For small molecule problems, several approaches have been pursued. At one extreme, finding the atoms to be superimposed of two or more small molecules reduces

to the maximal common substructure problem familiar from graph theory. Alternative strategies seek to identify an alignment compatible with what the receptor expects to see. Methods of this type include ASP,⁸ SQ,⁹ and FLEXS.¹⁰ Supervised alignment is typically embodied in a simple algorithm. The fitting of point sets is a general problem, and it has been tackled in a variety of different disciplines including multivariate statistical analysis, computer vision, and photogrammetry.¹¹⁻¹³

The following is a brief tutorial covering rotational superposition methods for vector sets, both in the molecular context and in others.

Parameterizing Rotation

Before this tutorial looks at rotational superposition, it is necessary to first look at the parameterizing of rotation. Rotation, translation, and reflection are the three most simple kinds of isometry. Any transformation of space that preserves distance is an isometry. Other isometries include the identity, central inversion, twist or screw displacement, glide reflection, and rotation inversion. Simple isometries, such as translation, combine in commutative pairs, to give compound isometries, such as glide reflection. Rotations are orthogonal transformations generating a matrix with determinant of +1. Matrices with determinant of -1 represent rotation inversion.

There are various different ways to parameterize rotation. Amongst the best known, and most applicable to coordinate transformation, are an arbitrary Eulerian rotation matrix, a polar-angle representation, and a quaternion-based representation. No 3-dimensional parameterization of rotation, such as the Eulerian matrix, is global and non-singular: there are points where the parameter values are not uniquely defined. 5-dimensions and above are required to represent rotation in a 1-1 global manner, although a 4-dimensional representation, such as quaternions, is sufficient for most practical purposes.

An Eulerian Rotation Matrix can be obtained by performing three rotations about three mutually perpendicular axes fixed in the body. We assume a right-handed frame for the axes, and define a positive rotation about a given axis. However, there is no unanimity in the choice of definition for Euler angles. For further discussion, see the Appendix.

Rotation around the X axis is given by:

$$R_x(U) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos U & \sin U \\ 0 & -\sin U & \cos U \end{bmatrix}$$

Rotation around the Y axis is given by:

$$R_y(U) = \begin{bmatrix} \cos U & 0 & -\sin U \\ 0 & 1 & 0 \\ \sin U & 0 & \cos U \end{bmatrix}$$

Rotation around the Z axis is given by:

$$R_z(U) = \begin{bmatrix} \cos U & \sin U & 0 \\ -\sin U & \cos U & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The rotation matrix is obtained by rotating these together in reverse order

$$R_{\phi\theta\theta} = R_i(\theta) R_{j\neq i}(\phi) R_{k\neq j}(\phi)$$

where subscripts relate to the axes chosen for rotation.

A polar angle representation is an alternative parameterization of rotation. Here, the rotation matrix is given by:

$$R_\lambda(U) = \begin{bmatrix} l(1-\cos U) + \cos U & ml(1-\cos U) + ns\sin U & nl(1-\cos U) - m\sin U \\ lm(1-\cos U) - ns\sin U & m^2(1-\cos U) + \cos U & mn(1-\cos U) + l\sin U \\ ln(1-\cos U) + ms\sin U & mn(1-\cos U) - l\sin U & n^2(1-\cos U) + \cos U \end{bmatrix}$$

where **n**, **m**, and **l** are components of the unit vector representing the rotation axis and **U** is the angle of rotation.

Finally, in this section of the tutorial, we turn to quaternions, which are also known as hypercomplex numbers, Rodrigues parameters, and, somewhat confusingly, Euler parameters. They are an ordered set of four numbers: one real and three imaginary components. Quaternion operators can rotate vectors about a given axis. The rotation matrix **R** is defined in terms of the quaternion parameters by the following:

$$R = \begin{bmatrix} 2(q_0^2 + q_1^2) - 1 & 2(q_0q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & 2(q_0^2 + q_2^2) - 1 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & 2(q_0^2 + q_3^2) - 1 \end{bmatrix}$$

where **q₀**, **q₁**, **q₂**, **q₃** are the components of a unit quaternion:

$$\begin{aligned} q_0 &= \cos(\theta/2) \\ q_i &= u_i \sin(\theta/2) \end{aligned}$$

which fulfills the condition:

$$\sum q_i^2 = 1$$

φ is the rotation angle and **u** is a normalized vector corresponding to the rotation axis. The determinant of this matrix reduces to:

$$D = (\sum q_i^2)^3$$

which is always equal to 1.0, whatever values are taken by the unit quaternion.

For the purposes of this tutorial, we can neglect other parameterizations of rotation. For discussion of these — the Cayley-Klein parameters, the Cayley parameterization, not to be confused with Cayley-Klein representation, unitary matrices, Pauli spin matrices, and many more (see references 14 and 15).

A Primer on the Superposition Problem

Before we discuss superposition algorithms, we shall continue this tutorial by considering the general case of vector set comparison. Having defined the equivalent atoms within a molecular structure, or indeed any multi-dimensional vector set, it is possible to determine their optimal fit in three dimensions. This process, usually known as superposition, is equivalent to finding the isometric transformation which best relates one set of vectors to another. A general rigid body transformation can be expressed as

the combination of a pure rotation and a translation. Thus the superposition of molecules reduces to finding the rotation and translation that best relates the two sets of atoms. Because these vector sets are seldom identical after transformation, it is necessary to formulate the fitting process in terms of an error function to be minimized; most usually as a least squares problem minimizing the discrepancy between equivalent co-ordinates. Expressing this in terms of vectors:

$$E_{TOTAL} = \sum (X - Y')^2$$

where Y' is the transformed vector set Y which best fits X . Y' is related to Y by

$$Y' = RY + t$$

where R is a rotation matrix and t a translation vector. Substituting the second equation into the first gives the error to be minimized in terms of rotation and translation:

$$E_{TOTAL} = \sum (X - (RY + t))^2$$

Differentiating E_{TOTAL} with respect to t , setting this expression to zero, and rearranging gives the following:

$$t = 1/n \sum X - 1/n \sum RY$$

It is clear from the above that the translation which best relates the two vector sets is the vector relating the respective centroids of the two sets. This is calculated easily and without knowledge of the required rotation. The best rotation relating the two vector sets is somewhat more difficult to determine and has received considerable attention. Typically, the best translation is found first, stored, and the co-ordinate sets translated to the origin.

Other, non-isometric comparisons of co-ordinate sets have been presented in the literature. Most obvious is to combine rotation and translation with a dilation:

$$Y' = RDY + t$$

where D is a dilation matrix. This matrix can be isotropic, applying a centrosymmetric scaling, or anisotropic. The optimal scaling factor can be obtained from

$$F_{x \rightarrow y} = \text{trace}[(R^T X^T)Y] / \text{trace}(X^T X)$$

where **trace** is the sum of the diagonal elements of a matrix. A more general least-squares formalism takes the following form:

$$Y' = DY + t$$

where

$$\begin{aligned} D &= RT \\ T &= (D^T D)^{1/2} \\ R &= D(D^T D)^{-1/2} \end{aligned}$$

Here, R is not the optimal rotation matrix for an orthogonal transformation but is a good starting point for iterative procedures. T is a real symmetric dilation matrix. Diagonalizing T , its eigenvectors show the principal strain vectors and its eigenvalues the magnitude of dilation. In a

famous but complex paper, Diamond has presented a method that adds a description of homogeneous strain and curvature to the analysis and develops quadratic transforms to compare structures.¹⁶ Likewise, Kearsley has also shown the utility of inhomogeneous transformation in structural comparison, focusing on cubic transforms and beyond.¹⁷ Lesk has recently investigated the use of Chebyshev fitting, where the error to be minimized is the maximum coordinate displacement and not the sum.¹⁸

Rotational Superposition of Two Vector Sets

Many different methods that find the non-linear least squares solution for the best rotation between two vector sets have been proposed. In this section of the tutorial, we shall consider such methods, and comment on the relative advantages and disadvantages of the algorithms. Such methods can be classified into two types: approximate or iterative methods and direct or closed-form methods. In a sense, this distinction is illusory, as most, if not all, direct methods require an eigenvalue determination which is typically an iterative procedure. However, the mathematical properties of these methods are distinct and their closed form affords the opportunity to analyze them rigorously.

The rotation minimizing the error between vectors in two dimensions has a simple closed-form solution. The rotation matrix R in two dimensions is given in terms of vectors by

$$R = \begin{bmatrix} \cos W & -\sin W \\ \sin W & \cos W \end{bmatrix}$$

where

$$W = \tan^{-1}(A)$$

and

$$A = \sum (X_1 X_2 - Y_1 Y_2) / \sum (X_1 X_2 + Y_1 Y_2)$$

over sets of atoms 1 and 2.

This forms the basis of an approximate, iterative superposition procedure used by a number of authors.¹⁹⁻²² Successive sets of rotations of the co-ordinates about the three orthogonal axes use this relation to best-fit equivalenced atoms. After several cycles, the process converges as the cumulative effect of matching in different mutually perpendicular planes reduces the overall change below a threshold. Dolata and Arnold have recently published a correction to a newly identified error in Nyburg's original algorithm.²³ Nyburg's method remains widely used. Other iterative methods^{24,25} are more computationally efficient as they require only a single rotation of co-ordinates.

Analytical, or closed-form, methods allow the direct calculation of the best transformation in a single step. Two classes of direct method may be discerned. The first makes use of the general properties of matrices, while the second uses an explicit representation, or parameterization, of rotation and the rotation matrix.

The key step of the majority of algorithms discussed below is the manipulation of a co-variance matrix of the matched atomic co-ordinates. This matrix, U , has the following form:

$$U = \begin{pmatrix} \sum X_1 X_2 & \sum X_1 Y_2 & \sum X_1 Z_2 \\ \sum Y_1 X_2 & \sum Y_1 Y_2 & \sum Y_1 Z_2 \\ \sum Z_1 X_2 & \sum Z_1 Y_2 & \sum Z_1 Z_2 \end{pmatrix}$$

over sets of atoms 1 and 2.

Of the first class of method, the algorithm of Kabsch is probably both the best known and the most widely used.^{26,27} This method requires an eigenvalue decomposition of the matrix $U^T U$. The resulting eigenvectors A^k are sorted and transformed (by application of the matrix U) to become the vectors B^k , and the best rotation matrix R is found from the expression:

$$R_{ij} = \sum_k B_i^k A_j^k$$

A method that also makes use of the covariance matrix U , but finds an analytical solution to the problem that involves complicated algebraic manipulations.²⁸ In practice, this can be dealt with better by iterative methods. An alternative method involves the eigenvalue decomposition of the partitioned matrix:²⁹

$$\begin{pmatrix} 0 & U \\ U & 0 \end{pmatrix}$$

A different approach that makes use of particular features of the problem to afford a fast solution using conjugate gradient optimization.³⁰ A similar method also makes use of techniques which optimize the initial properties of the system.²²

The most general approach of this type makes use of the singular value decomposition of U :

$$U = A D B$$

where A and B are 3x3 matrices and D is a diagonal matrix. The best rotation is obtained from:

$$R = B A'$$

Atomic coordinate matching^{29,31} is also known from work in other quite different disciplines, for example, in the literature of multivariate analysis,³²⁻³⁴ robotics, computer vision^{35,36} and numerical analysis.³⁷ The single value decomposition method is the most general approach to co-ordinate superposition since it is equally applicable to problems of higher or lower dimensionality.

To a greater or lesser extent, none of the analytical methods described above is completely general. All can give rise to rotation inversion (a rotation matrix with a determinant of -1) which inverts the chirality of a vector set. This can be corrected by multiplying the rotation inversion matrix by diag(1 1 -1). Certain of the methods are also sensitive to pathological conditions (such as the co-linearity or co-planarity of co-ordinates) which must be dealt with as special cases.

These failings are not shared by the second class of algorithm that makes use of an explicit formulation of the rotation matrix. Lesk presents a formulation of the problem that reduces to the maximization of a function of a single vari-

able, based on a parameterization of the rotation matrix in terms of polar angles.³¹ Liu and Van Rapenbusch published an explicit formulation of the superposition algorithm based around one form of the Eulerian matrix.³⁸ Other methods use an explicit representation of the rotation matrix based on the quaternion parameterization. MacKay was the first to propose such an approach in the molecular context.³⁹ However, his method minimizes the angle between vectors rather than the distance between points and thus is known not to generate an optimal superposition except in special cases.⁴⁰

Diamond presents a rigorous treatment in terms of quaternions which does generate an optimal rotation.⁴¹ Kearsley presents a simpler but more elegant algorithm using quaternions.⁴² All of the methods perform an eigenvalue decomposition of a 4x4 matrix, the resulting eigenvectors being the components of a unit quaternion (specifying the rotation), and the corresponding eigenvalue giving the value of the residual deviation after superposition. Kearsley's matrix has the following form:

$$\begin{pmatrix} \Sigma(X^2 + Y^2 + Z^2) & \Sigma(Z Y_+ - Y Z_+) & \Sigma(X Z_+ - Z X_+) & \Sigma(Y X_+ - X Y_+) \\ \Sigma(Z Y_+ - Y Z_+) & \Sigma(X^2 + Y_+^2 + Z_+^2) & \Sigma(X Y - Y_+ X_+) & \Sigma(Z X - X_+ Z_+) \\ \Sigma(Z X - X_+ Z) & \Sigma(X Y - Y_+ X_+) & \Sigma(X_+^2 + Y^2 + Z_+^2) & \Sigma(Z_+ Y - Y_+ Z_+) \\ \Sigma(Y_+ X_+ - Y X_+) & \Sigma(Z X_+ - X_+ Z) & \Sigma(Z_+ Y - Y_+ Z_+) & \Sigma(X_+^2 + Y_+^2 + Z_+^2) \end{pmatrix}$$

where

$$X_- = (X_1 - X_2) \text{ and } X_+ = (X_1 + X_2)$$

and the equivalent RMSD is given by

$$E_{mn} = (\lambda/n)^{1/2}$$

where n is the number of atoms fitted. The mathematical bases of Diamond's and Kearsley's methods are formally equivalent,⁴⁰ although their algorithmic implementation is different. Zuker and Somorjai present an alternative procedure that only uses information contained within the covariance matrix U .⁴³ Horn also presents a quaternion based method very similar to those discussed above.¹² Faugeras and Hebert present a similar result in the context of artificial intelligence research.⁴⁴ None of these methods give rise to special cases and, because of the quaternion form of the rotation matrix, cannot result in rotation inversion. Havel and Najfeld have derived a solution to the superposition problem using techniques from geometric algebra,⁴⁵ which closely parallels the quaternion treatment. Their method is both globally and quadratically convergent, and avoids sub-optimal solutions theoretically possible in certain quaternion formulations while converging more quickly.

Mairov and Crippen have examined random fits between proteins and shown that two structures are intrinsically similar if the mean-square deviation between them is less than when one structure is superposed to its mirror image.⁴⁶ This chiral molecular fitting can be obtained easily from quaternion-based fitting.⁴⁷ Using Kearsley's notation, the following inequality must be satisfied to allow chiral fitting:

$$r_2 + r_3 - r_1 - r_4 < 0$$

Quaternion, and indeed other methods, also allows antifitting where the RMSD can be maximized rather than minimized. Here the eigenvectors associated with the largest eigenvalue, rather than the smallest, are used to fit the structures. Using a quaternion-based approach, Salomon and Avnir have derived the best c_2 rotation for fitting a molecule to its own inverse.⁴⁸ They find the c_2 rotation axis as the eigenvector corresponding to the largest eigenvalue of the covariance matrix U.

Rotational Superposition of More Than Two Vector Sets

The superposition problem has been extended to cases of three or more co-ordinate sets. This is of obvious use in cases of protein structure comparison (such as comparison of NMR solutions) or the conformational analysis of small molecules (such conformational analysis). Unfortunately no closed-form solution exists for this problem, and so an iterative or optimization algorithm is required.

Haneef and co-workers developed a method in which an arbitrary number of sets were iteratively superposed onto a template or average structure, until convergence.⁴⁹ This is essentially, a re-discovery of a result originally presented by Gower.¹¹ Gerber and Muller and Shapiro et al. have both given more direct solutions based on matrix manipulations.^{50,51} Kearsley and Diamond have both presented methods based on quaternions.^{52,53} It is generally agreed that Diamond's method is the best available for multiple superimposition.

Sutcliffe and co-workers and Diamond have developed methods capable of clustering superposed co-ordinate sets into structurally related sets.^{54,55}

Software

There are many software implementations of rotational superposition. It is implemented as part of large commercial and semi-commercial software products such as SYBYL, O, or WHAT-IF. Rotational superposition is also incorporated into software that accomplishes unsupervised alignment.^{6,56}

Smaller, more dedicated programs also exist. These are all run on the host computer and are written in FORTRAN or C. I have written a program called FORFIT, which implements much of the above methodology. It can be downloaded from the home page of the Royal Society of Chemistry Molecular Modelling Group: <http://www.rsc.org/lap/rsccon/dab/ind006.htm>. Other programs include FIT, written by Guoguang Lu, which is available from <http://bilbo.bio.purdue.edu/~guoguang/fit.html> as FORTRAN source or executables. ProFit, written by Andrew Martin, is available from <http://www.biochem.ucl.ac.uk/~martin/swreg.html>, and LSQMAN, written by Gerard Klewegt, is available from <http://alpha2.bmc.uu.se/usf>. LsqKab, which implements Kabsch's methods^{26,27} is available from CCP4 (<http://www.dl.ac.uk/CCP/CCP4>). The program, Superimp, which implements Sippl's method²² is available from <ftp.came.sbg.ac.at/pub/SuperPos> as source code in C.

For multiple copy programs, PolyPose, which implements Diamond's method for multiple superposition, is

available from CCP4 (<http://www.dl.ac.uk/CCP/CCP4>). Program's, developed by Sutcliffe and co-workers, for multiple structural work are also available: NMRCLUST (<http://neon.chem.le.ac.uk/nmrclust/protocol.html>), and NMRCORE (<http://neon.chem.le.ac.uk/nmrcore/coreprot.html>). OLDERADO, a database created using these tools, is also available via the web (<http://neon.chem.le.ac.uk/olderado/index.html>). Of these programs, FIT and ProFit are written in C; the rest are written in FORTRAN.

Conclusions

In this tutorial, I have compared the nature of rotational superposition algorithms, mainly from a theoretical viewpoint. I have also compared the performance of methods published by Kabsch,^{26,27} Hendrickson,²⁵ McLachlan,^{29,30} and the quaternion based methods.^{12,41,42} This assessment included the accuracy of the obtained rotation matrix, the speed of the algorithm, and, where available, the returned RMS. All of the methods yield values of the rotation matrix which are correct and essentially the same, except for the iterative method of Hendrickson, which only gives optimal rotations when vector sets fit exactly. For this method, the discrepancy in the rotation matrix increases, as the fit between co-ordinate sets becomes poorer. As well as returning the best rotation, most of these methods return a value for the overall root mean squared deviation, in equivalence atoms, after superposition. Comparing these values to those measured from rotated co-ordinates suggests that some methods overestimate this deviation; typically the most accurate values are generated by quaternion methods.

Of available, and easily encoded, algorithms addressing pairwise solutions to the superposition problem, quaternion methods do not generate rotation inversions, are at least as fast as other methods, and produce more accurate RMS values. Since they have no special cases requiring separate treatment, the quaternion methods are easy to code. Despite the fact that quaternion methods are not well known,²³ and disagreements exist concerning their accuracy,⁴⁵ quaternion methods are the methods of choice for rotational superposition.

Appendix

Different formulations of the Eulerian Matrix

Twelve different formulations of the Eulerian matrix are given assuming anticlockwise rotations about mutually perpendicular axes in a right-handed co-ordinate system. $c1 = \cos \theta_1$ where θ_1 is the rotation about the first axes, $c2 = \cos \theta_2$ where θ_2 is the rotation about the second axes, etc. The key below gives the choice of axes for rotation, X = x axes, etc. The order refers to 1st, 2nd, and 3rd axes rotated around and not the order of matrices.

$$\text{KEY} = \begin{bmatrix} XYX & ZYX & XZX \\ YZX & YXY & ZXY \\ XZY & YZY & YXZ \\ ZXZ & XYZ & ZYZ \end{bmatrix}$$

Confusion arises, for example, from use of a left-handed co-ordinate system, clockwise rather than anticlockwise

$$\begin{pmatrix}
 c2 & s1s2 & -c1s2 \\
 s2s3 & c1c3 - s1c2s3 & s1c3 + c1c2s3 \\
 s2c3 & -c1s3 - s1c2c3 & -s1s3 + c1c2c3
 \end{pmatrix}
 \begin{pmatrix}
 c1c2 & s1c2 & -s2 \\
 -s1c3 + c1s2s3 & c1c3 + s1s2s3 & c2s3 \\
 s1s3 + c1s2c3 & -c1c3 + s1s2c3 & c2c3
 \end{pmatrix}
 \begin{pmatrix}
 c2 & c1s2 & s1s2 \\
 -s2c3 & c1c2c3 - s1s3 & s1c2c3 + c1s3 \\
 s2s3 & -c1c2s3 - s1c3 & c1c3 - s1c2s3
 \end{pmatrix}
 \\
 \begin{pmatrix}
 c1c2 & s2 & -s1c2 \\
 s1s3 - c1s2c3 & c2c3 & s1s2c3 + c1s3 \\
 s1c3 + c1s2s3 & -c2s3 & c1c3 - s1s2s3
 \end{pmatrix}
 \begin{pmatrix}
 c1c3 - s1c2s3 & s2s3 & -s1c3 - c1c2s3 \\
 s1s2 & c2 & c1s2 \\
 c1s3 + s1c2c3 & -s2c3 & -s1s3 + c1c2c3
 \end{pmatrix}
 \begin{pmatrix}
 c1c3 - s1s2s3 & s1c3 + c1s2s3 & -c2s3 \\
 -s1c2 & c1c2 & s2 \\
 c1s3 + s1s2c3 & s1s3 - c1s2c3 & c2c3
 \end{pmatrix}
 \\
 \begin{pmatrix}
 c2c3 & c1s2c3 + s1s3 & s1s2c3 - c1s3 \\
 -s2 & c1c2 & s1c2 \\
 c2s3 & c1s2s3 - s1c3 & s1s2s3 + c1c3
 \end{pmatrix}
 \begin{pmatrix}
 c1c2c3 - s1s3 & s2c3 & -s1c2c3 - c1s3 \\
 -c1s2 & c2 & s1s2 \\
 c1c2s3 + s1c3 & s2s3 & -s1c2s3 + c1c3
 \end{pmatrix}
 \begin{pmatrix}
 c1c3 + s1s2s3 & c2s3 & -s1c3 + c1s2s3 \\
 -c1s3 + s1s2c3 & c2c3 & s1s3 + c1s2c3 \\
 s1c2 & -s2 & c1c2
 \end{pmatrix}
 \\
 \begin{pmatrix}
 c1c3 - s1c2s3 & s1c3 + c1c2s3 & s2s3 \\
 -c1s3 - s1c2c3 & -s1s3 + c1c2c3 & s2c3 \\
 s1s2 & -c1s2 & c2
 \end{pmatrix}
 \begin{pmatrix}
 c2c3 & s1s2c3 + c1s3 & -c1s2c3 + s1s3 \\
 -c2s3 & -s1s2s3 + c1c3 & s3s2c1 + s1c3 \\
 s2 & -c2s1 & c1c2
 \end{pmatrix}
 \begin{pmatrix}
 c1c2c3 - s1s3 & s1c2c3 + c1s3 & -s2c3 \\
 -c1c2s3 - s1c3 & -s1c2s3 + c1c3 & s2s3 \\
 c1s2 & s1s2 & c2
 \end{pmatrix}$$

rotations, or, as Euler did himself, a combination of clockwise and anticlockwise rotations. ZXZ is probably the most widely used convention, as many physicists in their treatment of mechanics have adopted it. Rossman and Blow, in their famous formulation of molecular replacement,⁵⁷ use a similar convention, although many other conventions have also been adopted in other formulations of molecular replacement.⁵⁸ The angles of the XYZ convention are also called Bryant or Cardan angles. The aeronautical, automotive, and nautical engineers have adopted the ZYX convention, with its interpretation of yaw, pitch, and roll. Likewise, the YXZ convention can be interpreted as pan, tilt, and cant. Other physicists have adopted ZYZ as their convention.

References

- Marriott, D.P., Dougall, I.G., Meghani, P., Liu, Y.-J., and Flower, D.R. 1999. Lead generation using pharmacophore mapping and 3D database searching. Application to muscarinic M3 antagonists. *J. Med. Chem.* 42:3210-3216.
- Cramer, R.D., Patterson, D.E., Bunce, J.D. 1988. Comparative Molecular-Field Analysis (CoMFA) I. Effect of shape on binding of steroids to carrier proteins. *J. Amer. Chem. Soc.* 110:5959-5967.
- Holm, L. and Sander, C. 1994. Searching protein-structure databases has come of Age. *Proteins-Structure Function and Genetics* 19:165-173.
- Flower, D.R. (1993) Structural relationship of streptavidin to the calycin protein superfamily. *FEBS Lett.* 333:99-102.
- Flower, D.R., North, A.C.T., and Attwood, T.K. 1993. Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci.* 2:753-761.
- Chothia, C. and Lesk, A.M. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5:823-826.
- Brown, N.P., Orengo, C.A., and Taylor, W.R. 1996. A protein structure comparison methodology. *Computers & Chemistry* 20:359-380.
- Hodgkin, E.E. and Richards, W.G. 1987. Molecular Similarity based on electrostatic potential and electric-field. *International J. Quantum Chem.* 14:105-110.
- Miller, M.D., Sheridan, R.P., and Kearsley, S.K. 1999. Sq: a program for rapidly producing pharmacophorically relevant molecular superpositions. *J. Med. Chem.* 42:1505-1514.
- Lemmen, C., Lengauer, T., and Klebe, G. 1998. FlexS: a method for fast flexible ligand superposition. *J. Med. Chem.* 41:4502-4520.
- Gower, J.C. 1975. Generalized procrustes analysis. *Psychometrika* 40:33-51.
- Horn, B.K.P. 1987. Closed-form solution of absolute orientation using unit quaternions. *J. Optical Soc. Amer. A* 4:629-642.
- Sabata, B. and Aggarwal, J.K. 1991. Estimation of motion from a pair of range images — a review. *CVGIP-Image Understanding* 54:309-324.
- Rooney, J. 1977. A Survey of representations of spatial rotation about a fixed point. *Environment and Planning B* 4:185-210.
- Stuepnagel, J. 1964. On the parameterisation of the three-dimensional rotation group. *SIAM Rev.* 6:422-430.
- Diamond, R. 1976. On the comparison of conformations using linear and quadratic transformations. *Acta Cryst. A* 32:1-10.
- Kearsley, S.K. 1989. Structural comparisons using restrained inhomogeneous transformations. *Acta Cryst. A* 45:628-635.
- Lesk, A.M. 1998. Extraction of geometrically similar substructures: least-squares and Chebyshev fitting and the difference distance matrix. *Proteins-Structure Function and Genetics* 33:320-328.
- Nyburg, S.C. 1974. A best rigid-body molecular fit routine. *Acta Cryst. B* 30:251-253.
- Kenknight, C.E. 1984. Comparison of methods of matching protein structures. *Acta Cryst. A* 40:708-712.
- Horn, B.K.P., Hilden, H.M., and Negahdaripour, S. 1988. Closed-form solution of absolute orientation using orthonormal matrices. *J. Optical Soc. Amer. A* 5:1127-1135.
- Sippl, M.J. and Stegbuchner, H. 1991. Superposition of three-dimensional objects: a fast and numerically stable algorithm for the calculation of the matrix of optimal rotation. *Comp. & Chem.* 15:73-78.
- Dolata, D.P. and Arnold, J. 1992. An easily fixed error in the Nyburg algorithm for discovering the best fit between molecules. *J. Mol. Graph.* 10:222-224.
- Ferro, D.R. and Hermans, J. 1977. A different best fit rigid-body fit routine. *Acta Cryst. A* 33:345-347.
- Hendrickson, W. 1979. Transformations to optimise the superposition of similar structures. *Acta Cryst. A* 35:158-163.
- Kabsch, W. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Cryst. A* 32:92-933.

27. Kabsch, W. 1978. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst.* A34, 827-828.
28. McLachlan, A.D. 1972. A mathematical procedure for superimposing atomic co-ordinates of proteins. *Acta Cryst.* A28:656-657.
29. McLachlan, A.D. 1979. Least squares fitting of two structures. *J. Mol. Biol.* 128:74-77.
30. McLachlan, A.D. 1982. Rapid Comparison of Protein Structures. *Acta Cryst.* A38:871-873.
31. Lesk, A.M. 1986. A toolkit for computational molecular-biology 2. On the optimal superposition of 2 sets of coordinates. *Acta Cryst.* A42:110-113.
32. Cliff, N. 1966. Orthogonal rotation to Congruence. *Psychometrika* 31:33-42.
33. Schonemann, P.H. 1966. A generalised solution of the orthogonal procrustes problem. *Psychometrika* 31:1-10.
34. Schonemann, P.H. and Carroll, R.M. 1970. Fitting one matrix to another under a choice of a central dilation and a rigid motion. *Psychometrika* 35:245-255.
35. Arun, K.S., Huang, T.S., and Blostein, S.D. 1987. Least-squares fitting of two 3-D point sets. *I.E.E.E. Transactions On Pattern Analysis and Machine Intelligence* 9:699-700.
36. Umeyama, S. 1991. Least-squares estimation of transformation parameters between two point patterns. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence* 13:376-380.
37. Golub, G. and Loan, C.V. *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1989.
38. Liu, Z. J. & Van rapenbusch, R. 1988. A fast, direct algorithm for the least-squares fitting of two sets of atomic coordinates of macromolecular structures. *J. Comp. Chem.* 9:596-599.
39. Mackay, A.L. 1984. Quaternion transformation of molecular-orientation. *Acta Cryst.* A40:165-166.
40. Diamond, R. 1989. A comparison of three recently published methods for superimposing vector sets by pure rotation. *Acta Cryst.* A45:657.
41. Diamond, R. 1988. A note on the rotational superposition problem. *Acta Cryst.* A44:211-216.
42. Kearsley, S.K. 1989. On the orthogonal transformation used for structural comparisons. *Acta Cryst.* A45:208-210.
43. Zuker, M. and Somorjai, R.L. 1989. The alignment of protein structures in 3-dimensions. *Bull. Math. Biol.* 51:55-77.
44. Faugeras, O.D. and Hebert, M. 1983. Proceedings of the eighth international joint conference on artificial intelligence, p. 996-100. Los Altos, CA: William Kaufmann.
45. Havel, T.F. and Najfeld, I. 1994. Applications of geometric algebra to the theory of molecular conformation. Part 1. The optimum alignment problem. *J. Mol. Struct. (THEOCHEM)* 308:241-262.
46. Maiorov, V.N., Crippen, G.M. 1994. Significance of root-mean-square deviation in comparing 3-dimensional structures of globular-proteins. *J. Mol. Biol.* 235:625-634.
47. Diamond, R. 1990. Chirality in rotational superposition. *Acta Cryst.* A46:423.
48. Salomon, Y. and Avnir, D. 1999. Continuous symmetry measures: finding the closest C₂-symmetric object or closest reflection-symmetric object using unit quaternions. *J. Comp. Chem.* 20:772-780.
49. Sutcliffe, M.J., Haneef, I., Carney, D., and Blundell, T.L. 1987. Knowledge-based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.* 1:377-384.
50. Gerber, P.R and Muller, K. 1987. Superimposing several sets of atomic coordinates. *Acta Cryst.* A43:426-428.
51. Shapiro, A., Botha, J.D., Pastore, A., and Lesk, A.M. 1992. A method for multiple superposition of structures. *Acta Cryst.* A48:11-14.
52. Kearsley, S.K. 1990. An algorithm for the simultaneous superposition of a structural series. *J. Comp. Chem.* 11:1187-1192.
53. Diamond, R. 1992. On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci.* 1:1279-1287.
54. Kelley, L.A., Gardner, S.P., and Sutcliffe, M.J. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* 9:1063-1065.
55. Diamond, R. 1995. Coordinate-based cluster-analysis. *Acta Cryst.* D51:127-135.
56. May, A.C.W. 1996. Pairwise iterative superposition of distantly related proteins and assessment of the significance of 3-D structural similarity. *Protein Eng.* 9:1093-1101.
57. Rossman, M.G. and Blow, D.M. 1962. The detection of subunits within the crystallographic assymmetric unit. *Acta Cryst.* 15:24-30.
58. Urzhumtseva, L.M. and Urzhumtsev, A.G. 1997. Tcl/Tk-based programs 2. Convrot: a program to recalculate different Rotation descriptions. *J. Appl. Cryst.* 30:402-410.

Teaching Computational Chemistry in the Undergraduate and Graduate Chemistry Curriculum*

Thomas Holme

Department of Chemistry, University of Wisconsin – Milwaukee, P.O. Box 413, Milwaukee, WI 53201. 414-229-3970. tholme@uwm.edu

I present an example of teaching an advanced undergraduate course in computational chemistry. The primary goals of the course are to provide practical experience using various computational chemistry methods, basic theoretical background on what methods apply to problems, and an awareness of the widespread use of computational techniques in modern chemistry research.

Introduction

As computational chemistry becomes increasingly incorporated into mainstream research efforts of non-computa-

*Based on a paper presented (abstract COMP 172) at the 218th National Meeting of the American Chemical Society, New Orleans, LA, August 22-26, 1999.

tional groups, the need for enhanced teaching in the field becomes more evident. As with any research tool, when applied improperly computational results can be grossly misleading. Now that the challenging work of making computational chemistry widely available has been largely accomplished, it only makes sense that the dedicated practitioners of computational chemistry educate the uninitiated.

From an educational perspective, those who emphasize curriculum and materials development have tended to view the importance of computers in chemistry education from either a tutorial perspective¹ or as a means to better visualize chemicals or chemical reactions.^{2,3} Visualization, in particular, has received considerable attention with the advent of graphics capability⁴ alongside scattered, individual efforts to apply visualization in various components of the chemistry or science curriculum.⁵ Moreover, cognitive science research has investigated the impact computer visualization has on student learning⁶ and found some measurable positive results. In addition to these formal literature reports of pedagogical interest in computational chemistry, symposia have been organized at most national meetings of the American Chemical Society where computer applications in chemistry have been discussed.

Unfortunately, emphasis on visualization tends to treat the required computation, and the model, as a black box, and thus there remains a need for courses where students learn more completely how computational chemistry is accomplished. Two trends are emerging for accomplishing this goal. One method is to integrate computational chemistry throughout the undergraduate major in chemistry. This approach has been reported at both University of North Carolina, Wilmington⁷ and California State University, Fullerton.⁸ In this scenario, an introductory course is provided early in the curriculum to facilitate use of computational methods in most other undergraduate courses. Because the course is offered before the physical chemistry course, the theory base that may be covered is constrained. Nonetheless, such a course is capable of providing important instruction, particularly when considered within the context of an integration of computational chemistry in most undergraduate courses.

A second approach is discussed here and has been reported from the University of Chicago.⁹ In this case, computational concepts and practice are taught as a separate course taken by advanced undergraduate students and entering graduate students. Because the course is offered for one semester or quarter, the amount of material that can be included is limited. I present this course as one example of how computational chemistry might be organized into an upper division undergraduate course.

Goals

We identified three primary goals for our new course. The first goal is to provide students with practical exposure to computational chemistry methods. Hands-on experience with a variety of important methods is an invaluable learning tool and simply must be accommodated. Second, students need exposure to a basic level of theory in order to make informed choices about the possibilities and limita-

tions of various computational methods. This theoretical background is motivated by the need to keep students from using methods to answer questions these methods are not equipped to address. Third, students should be made aware of the rapidly expanding research applications of computational chemistry. They will not know the possibilities if they have not observed how the methods are currently being used. In particular, because few of our students are interested in doctoral research in computation chemistry (the course is in a service course) they need to find out how their own research interests concur with the utility of computational methods. We address this goal using a current research literature component for the course – one that we will not discuss in this paper because it is highly variable based on both student interests and trends in the literature.

Structure

The course is a three-credit, one-semester course formally offered from the physical chemistry division of our department. The class meets for three 50 minute periods — though time spent in hands-on work with software often extends beyond the allotted time. One period is a lecture, another is hands-on work, and a third is working on current literature. We have had no required textbook, though we encourage students to obtain a book on basic molecular orbital theory¹⁰ and on computational methods.¹¹ We believe that newer alternatives for this latter book will be used when the course is offered again.¹²

Lectures

The lecture component of the course is designed to provide enough background that students will use computational chemistry in an informed manner. With this intent in mind, we seek to limit the possibilities that improper calculations will be carried out by students by informing them about the possibilities and limitations of several critical methods. The first lecture serves as an overall summary of many computational techniques — essentially to identify methods that will not be emphasized but are, nonetheless, important research tools. Such subjects include combinatorial chemistry and databases, reaction dynamics, molecular dynamics, and Monte Carlo techniques.

Beginning from the second lecture, we emphasize structural determinations through computational chemistry. This choice is largely motivated by our student clientele — many of whom have joined research groups who use computational chemistry for this type of study. We begin, therefore, with molecular mechanics (MM) techniques. The type of topics covered in these lectures include the essential structure of MM force fields; how the use of force fields thus defined constrains the type of questions that can be asked; what are the best questions to ask using MM techniques; and how does one develop reliable force fields for new functional groups of interest. Interspersed throughout this theoretical development are regular references to specific MM implementations using both the Tripos force field¹³ and MM3.¹⁴ More pragmatic topics such as torsional drivers and minimization techniques are also introduced.

Our next goal in the lecture portion of the course is to

introduce quantum mechanical techniques and software. We carry out a survey of student understanding of molecular orbital theory prior to reaching this point in our course. In our experience, students need some remedial work with MO theory prior to covering any details about ab initio methods for quantum mechanical calculations. In our case, we believe that reestablishing an intuitive feel for MO theory with schematic representations is a most valuable starting point. We therefore cover molecular orbital (MO) theory at a level commonly found in senior level inorganic chemistry courses. Once we have this intuitive picture firmly established, we proceed to describe computational implementations of MO theory. Specifically we cover the basic theory associated with Hartree-Fock self-consistent field (HF-SCF) calculations such that students are aware of the fundamental assumptions in such calculations. They also learn the differences between semi-empirical approaches and ab initio ones. Inherent in these discussions are such items as basis sets, canonical orbitals, and localization of orbitals, SCF conditions, convergence criteria, energy gradients, and restricted versus unrestricted Hartree-Fock. We then consider methods for addressing electron-electron correlation with emphasis on perturbation theory and multi-configuration SCF (MCSCF).

The final topic introduced in the lecture portion of the course is density functional theory. Unfortunately, this topic receives less coverage because we tend to run out of time at the end of the semester.

Some sample questions from the final exam provide insight into the level of expectation we have for the students in our course: (i) What is the basic approximation for electron-electron interactions in Hartree-Fock theory?; (ii) What chemical questions can be answered using molecular mechanics calculations? What cannot be answered with them?; (iii) What is the difference between a semi-empirical calculation and an ab initio one? (consider both methodology and computer resource utilization); and (iv) What is the difference between a primitive and a basis function?

Hands-on Work

Our goal in the hands-on sessions was to provide opportunities for students to work with a number of methods and then allow them to apply appropriate tools to a question of their own design. The majority of students in our course are entering graduate students who have some idea of research interests so they are quite capable of devising their own project, which is the assignment for the last several weeks of the course. Our laboratory for students consists of Sun Ultra 30 workstations. The students learn to use the SYBYL¹⁵ package as a user interface, and then learn to use MM3¹⁴ and GAMESS.¹⁶

Our first introduction to SYBYL involves drawing four conformers of glycine and determining which has the lowest energy. Students are then required to comment on why their answer makes sense. The second assignment is to choose any three amino acids and determine the lowest energy primary structure for the tri-peptide. Although a full conformational search is not required, the students are expect-

ed to discuss the sensibility of their result. The third assignment is a z-matrix assignment. This assignment was designed to prepare students to use the GAMESS input scheme. The fourth assignment was to use AM1¹⁷ calculations to determine the lowest energy isomer of a series of bicyclic compounds. Again, students are required to comment on the reasonability of their calculated results. The fifth assignment is designed to have students looking at molecular orbitals that result from calculations on various diatomic molecules and compare them with those they predict using cartoon type MO diagrams. This assignment also focuses on issues associated with the HOMO-LUMO gap. The sixth assignment takes the issues of orbitals further: using them to predict the location of nucleophilic attack. In this case, the students find that the computer visualization is more difficult to interpret than looking at the eigenvectors of the orbitals involved. We are thus able to nudge students away from looking at the computational package as a black box. The final set assignment was to geometry optimize and characterize the energy minimum structure for acetic acid using both semi-empirical calculations and small basis set ab initio calculations. The normal mode vibrations were assigned by the students and compared with literature values. The final five weeks of the semester, we devoted to having the students carry out a project of their own choosing. The project had as a minimum requirement the determination of at least one local minimum and one transition state, but any system could provide the molecules of interest. This assignment, in a class of 12 students, required significant additional instructional time.

Student Response

Student response to the course has been generally positive. Many students would prefer more time with the instructor in the laboratory. Few students could finish the projects assigned in the allotted 50 minutes. The library portion of the course, despite being catered quite strongly to the individual students received mixed reviews. We, nonetheless, feel the requirement of library research is an important component of how students can learn about the utility of computational chemistry. Many of the concerns about this component were associated with the heavy writing load it induced in compiling reports. As for the theory content, students tend to perceive it as appropriate or perhaps slightly too little. We do not intend to make any major changes in the theory content in the future, though the availability of good textbooks may alter this expectation.

In summary, our experience with the teaching of a one-semester computational chemistry course has been largely positive. There will remain room for fine-tuning of this course, and it is clearly not the only way to introduce this field into the chemistry curriculum. The primary benefit of including such a course lies in having students who are in experimental research groups make good decisions about ways that computational chemistry can be used. In our experience, the effort of developing this type of course is energy well spent.

References

1. Appling, J.R. and Frank, D.L. Discover Chemistry 2.0. Brooks/Cole, Pacific Grove, CA, 1998.
2. Beckwith, E.K., and Nelson, C. The ChemViz project: Using a supercomputer to illustrate abstract concepts in chemistry. Learning and Leading with Technology. 1998, 25:17-19.
3. Klassen, J.B., Graham, K.J., and Muldoon, W.P. Molecular modeling as an aid to understanding stereoselectivity. *J. Chem. Educ.* 1999, 76:985-986.
4. Santa Coloma, T.A. and Calvo, J.C. The use of computers in the teaching of hormone receptor interactions in the presence of two types of binding sites or negative cooperativity. *Biochem. Educ.*, 1998, 16:90-91. Elgren, T.E. Consideration of Lewis acidity in the context of heme biochemistry: A molecular visualization exercise. *Chemical Educator* s1430-4171 (98) 03206-7 <http://chedr.edoc.com/chedr/>
5. Lewis, N.S. The computer animation project. <http://bond.caltech.edu>
6. DeKock, R.L., Madura, J.D., Rioux, F., and Casanova, J. Computational chemistry in the undergraduate curriculum. In: *Reviews in computational chemistry*, Lipkowitz, K.B., and Boyd, D.B., Eds., VCH Publishers, New York, 1993, Vol. 4, pp. 149-228. Lipkowitz, K.B., Jalaie, M., Robertson, D., and Barth, A. Interdisciplinary learning with computational chemistry: a collaboration between chemistry and geology. *J. Chem. Educ.* 1999, 76:684-686
7. Williamson, V.M. and Abraham, M.R. The effects of computer animation in the particulate mental models of college chemistry. *J. Res. Science Teaching* 1995, 32:521-534. Cracolice, M.S., and Abraham, M.R. A comparison of computer-assisted instruction, semi-programmed instruction, and teaching assistant-led instruction in general chemistry. *School Science and Mathematics* 1996, 96:215-221.
8. Martin, N.H. Integration of computational chemistry into the chemistry curriculum. *J. Chem. Educ.* 1998, 75:241-243.
9. Kantardjieff, K. Enhancing instruction in upper division chemistry at California State University Fullerton. *Chemical Educator* S1430-4171 (97) 02112-2 available at <http://chedr.edoc.com/chedr/>. Kantardjieff, K., Hardinger, S.A. and Van-Willis, W. Introducing computers early in the undergraduate chemistry curriculum. *J. Chem. Educ.* 1999, 76:694-696.
10. Gasyna, Z.L., and Rice, S.A. Computational chemistry in the undergraduate chemistry curriculum: development of a comprehensive course formula. *J. Chem. Educ.* 1999, 76:1023-1029.
11. Clark, T. A. *Handbook of computational chemistry*. Wiley, New York, 1985.
12. Goodman, J. *Chemical Applications of Molecular Modelling*. Royal Society Chemistry, London, 1998. Grant, G.H. and Richards, W.G. *Computational Chemistry*. Oxford Univ. Press, New York, 1995. Leach, A.R. *Molecular Modelling: Principles and applications*. Addison Wesley Longman, New York, 1996.
13. Clark, M., Cramer, R.D. III, and Van Opdenbosch, N. Verification of the general purpose Tripos 5.2 force field. *J. Comp. Chem.* 1989, 10:982-1012.
14. Allinger, N.L., Zhou, X. and Bergsma, J. Molecular mechanics parameters. *J. Mol. Struct. (THEOCHEM)* 1994, 312:69-83. Pettersson L. and Liljefors, T. Molecular mechanics calculated conformational energies of organic molecules: a comparison of force fields. In: *Reviews in computational chemistry*, Lipkowitz K.B. and Boyd, D.B., Eds., VCH Publishers, New York, 1996, 9:167-189.
15. SYBYL, version 6.6, 1999, Tripos Inc., St. Louis, MO.
16. Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.J., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., Windus, T.L., Dupuis, M., and Montgomery, J.A. General atomic and molecular structure system. *J. Comp. Chem.* 1993, 14:1347-1363.
17. Stewart, J.J.P. Special Issue on MOPAC — A Semiempirical Molecular-Orbital Program. *J. Comp.-Aided Mol. Design.* 1990, 4:1-105.

COMPANY PROFILE**OpenEye Scientific Software, Inc.**

Anthony Nicholls, President, OpenEye Scientific Software, Inc., 335c, Winische Way, Santa Fe, NM 87501, USA 505-982-6354. www.eyesopen.com

The amortized computer cost of construction of a multi-conformer library of a typical pharmaceutical corporate collection is roughly 30 cents. The cost of disk storage is another dollar, or 20 cents on a CD. Contrast this with typical molecular modeling projects on tens to hundreds of compounds and it becomes clear that the astonishing progress of computer hardware is not being adequately harnessed to lead discovery in the pharmaceutical industry.

OpenEye Scientific Software was founded in 1996, its goal to provide modelers with new tools for the analysis of the 3D properties of molecules. Rather than focus on the traditional modeling areas, such as force-fields and quantum mechanics, these tools concern molecular shape and electrostatics. Innovations in both make OpenEye tools extensible to library construction on a scale commensurate with today's hardware, i.e., at a scale currently unenvisioned.

OpenEye's first product, OMEGA, a conformer ensemble generator, can process over half a million drug-like molecules per day per typical processor, making virtual library construction straightforward. Structures are generated by torsion search streamlined by rules that can be user-modified to include chemical intuition or information. Diversity is ensured by r.m.s. culling. Protein structures can guide library construction via active-site templates and the position of three or more ligand atoms. Conformers require only about twenty bytes to store, speeding disk and in-memory access. OMEGA is available for trial.

Drug interactions are aqueous, yet most molecular modeling still uses vacuum force-fields. OpenEye uses ZAP, an innovative Poisson-Boltzmann solver, to model water. Conformations can be minimized and ranked against solvent rather than vacuum energies. The electrostatic component of drug-protein interactions can be rapidly calculated. Solvent-mediated *pKa* shifts can be estimated in drugs and protein active sites. QSCORE, a ZAP-derived procedure, allows grid-based estimation of desolvation energies of thousands of drug-protein orientations per second.

OpenEye's ESPACE approach decodes the shape of a structure into typically three or four disjoint ellipsoids, fit to the molecular volume. Global and partial shape matches become possible in times previously reserved for sub-

UPCOMING MEETINGS

structure search. Properties such as hydrogen-bond donor/acceptor propensity can be associated with the surface of each ellipsoid and used as an additional filter. ESEARCH hits are returned with a rigorously defined shape and shape-property Tanimoto. Ellipsoidal representations of libraries (ELIBS) allow for novel descriptors of diversity and also the comparison of collections free of proprietary information.

All OpenEye code is C/C++ and has been ported to most major platforms, including, and especially, Linux. We believe our approach, coupled with ever-cheaper storage and computation, will change the possibilities of molecular modeling.

UPCOMING MEETINGS

February 2000

High-Throughput Organic Synthesis

February 10-11, 2000

Hyatt Regency La Jolla
3777 La Jolla Village Drive
San Diego CA 92122, USA

Contact

Jennifer Laakso
Cambridge Healthtech Institute
1037 Chestnut Street
Newton Upper Falls MA 02464 USA
Tel.: 617-630-1300
Fax: 617-630-1325
chi@healthtech.com
www.healthtech.com

2000 Charleston Conference: Innovative Techniques for Lead Discovery and Development

February 28- March 1, 2000

Wild Dunes Resort
Isle of Palms, SC USA

Over the past few years, competitive pressures have forced research to become more efficient. Emerging new technologies, such as genomics, gene sequencing, transgenic animals, and molecular biology, have afforded large numbers of novel, clinically relevant biological targets. High throughput screening and combinatorial chemistry have been used to rapidly identify and optimize leads for these targets. Additionally, the complementary techniques of computational chemistry and cheminformatics have been used to successfully leverage research information.

The invited lectures, discussions, and poster session to be held over three days will focus on the successful integration of these technologies in new lead discovery and development. Participants will explore the ways in which the industry is increasing the productivity of the discovery process.

Contact

Merry Ambos
Network Science Corporation
412 Carolina Blvd.
Isle of Palms, SC 29451-2113 USA
Tel.: 843-886-5008, Fax: 843-886-5924
mambos@netsci.org
www.netsci.org

March 2000

2000 Charleston Conference: Research Perspectives in Computational and Structural Chemistry

March 1-3, 2000

Wild Dunes Resort
Isle of Palms, SC USA

The complementary techniques of computational chemistry and informatics have been used to successfully leverage research information which has contributed to the discovery and optimization of promising new chemical entities. When combined with evolving technologies, such as genomics, proteomics, ultra high throughput screening, and combinatorial synthesis, computational protocols, such as QSAR, virtual screening, docking, and diversity analysis, can be used to direct discovery and optimization.

The invited lectures, discussions, and poster session to be held over three days will focus on the successful integration of these technologies in new lead discovery and development. Participants will explore the ways in which the industry is increasing the productivity of the discovery process.

Contact

Merry Ambos
Network Science Corporation
412 Carolina Blvd.
Isle of Palms, SC 29451-2113 USA
Tel.: 843-886-5008, Fax: 843-886-5924
mambos@netsci.org
www.netsci.org

Second Log P Symposium. Lipophilicity in Drug Disposition: Practical and Computational Approaches to Molecular Properties Related to Drug Permeation, Disposition, and Metabolism

March 5-9, 2000

University of Lausanne, Switzerland

The Symposium will focus on the determination, computation, and interpretation of lipophilicity and related molecular properties as factors and predictors of drug permeation, disposition, and metabolism. Strong introductory emphasis will be given to the biological and pharmacokinetic background.

In experimental themes, particular attention will be paid

to the lipophilicity profiles of ionized compounds, to lipophilicity measurements in anisotropic media (liposomes/water, IAM columns), and to permeability across artificial membranes. The relevance of these parameters to pharmacokinetic properties will be examined. Computational themes will comprise lipophilicity and H-bonding fields and their interest in docking strategies and structure-permeation relations.

Throughout the Symposium, lipophilicity and related parameters will be contemplated from a dual perspective, their interpretation in terms of recognition forces, and their value in screening, lead optimization, and drug candidate selection.

Scientific and Organizing Committee

- Prof. Bernard Testa, Chair (School of Pharmacy, University of Lausanne)
- Prof. Gerd Folkers (Dept of Pharmacy, Swiss Federal Institute of Technology, Zurich)
- Dr. Pierre-Alain Carrupt (School of Pharmacy, University of Lausanne)
- Mrs. Nicole Matter (Lausanne Tourism)
- Dr. Joachim Mayer (School of Pharmacy, University of Lausanne)
- Dr. Marianne Reist (School of Pharmacy, University of Lausanne)
- Dr. Han van de Waterbeemd (Pfizer Central Research, Sandwich, UK)

Contact

Register online at <http://ictsg10.unil.ch/logp2000>

BIO2000

March 26-30, 2000

Hynes Convention Center, Boston, MA

This conference covers all aspects of biotechnology, especially business. Two of the symposia are listed below:

Instant Information: How the Internet is Changing Biotechnology

The explosive growth and far-reaching influence of the World Wide Web and the Internet, communications are rapidly changing every aspect of healthcare and biotechnology. Millions of adults worldwide will soon access the Internet regularly for information on scientific discoveries, medical research, new products, investment opportunities, and other healthcare consumer resources. This symposium will address each of these components, from online publishing and banking to clinical trial recruitment and Web-based patient databases.

Pharmacogenomics - Realizing the Full Potential of the Human Genome

Everyone is wondering when the Human Genome Project will yield its promise in terms of better patient care. The discussion will address the implications for medical practice: Will genetic testing become a more lucrative field? Will prophylactic drugs get a boost? How will the industry deal with personalized medicine?

Contact

Register online at <http://www.bio.org/events/2000/bio2000.html>

GENOMICS: New Discoveries and Commercial Developments

March 29-31, 2000

Churchill College, Cambridge, Cambridgeshire, UK

Genomics is poised to play an increasing role in medicine and agriculture. In the next few years, the complete sequences of hundreds of genomes will be available and recent developments have indicated that the human genome sequence will be finished ahead of the 2005 target date. These sequence data will be accompanied with a vast amount of additional information including results of comparative genomics, RNA and protein expression profiles, functional and pathway information, and tools for molecular and cell biology.

How is the wealth of genome information going to be interpreted? How is this information going to be managed and shared? What impact has genomics had in the commercial environment and what are the future opportunities for exploitation of genomic information? This conference will provide a broad, in depth overview of new discoveries and commercial developments in genomics and will provide a framework for timely discussion of key issues.

The program has been designed to appeal to scientists working in the area of genomics, to those developing enabling technologies, and to managers considering broader application of genomics in the Pharmaceutical and Agribusiness industries.

Session	Included Topics
Structural Genomics	Progress in plant and animal sequencing Future Challenges
Gene Function & Expression	Expression profiling Antisense Gene Regulation Pathways
Bioinformatics	Data mining and visualization Genome mapping
Applications & Opportunities	Gene expression Cells as factories Pharmacogenetics
The Way Forward?	Commercial developments Impact on the industries

Contact

Sonia Walter
Society of Chemical Industry
14/15 Belgrave Square
London SW1X 8PS UK
Tel.: +44 (0) 171 235 3681
Fax: +44 (0) 171 235 7743
soniaw@chemind.demon.co.uk
<http://sci.mond.org/>

April 2000**Mathematical and Statistical Aspects of Molecular Biology, 10th Annual Meeting***April 4-5, 2000**Isaac Newton Institute for Mathematical Sciences,
Cambridge, UK*

The meeting aims to bring together mathematicians, statisticians, biologists, and computer scientists who have common research interests in the analysis of molecular biological data.

The MASAMB meetings generally consist of talks and discussions on the theory of methods used for the analysis of molecular sequence data. Applications of these methods are also of interest, but generally more for the light they throw on the methods than for their biological conclusions.

Included Topics

- Molecular phylogenetics (sequence evolution, phylogenetic inference)
- Population genetics (coalescent-based methodology, analysis of mutation data, forensic science)
- Genetic mapping and linkage
- Gene finding and motif detection
- Sequence database searching and alignment
- Protein structure prediction and analysis

Contact

Scientific Organizer: Nick Goldman (Department of Genetics, University of Cambridge)
<http://www.newton.cam.ac.uk/masamb.html>

**18th International MGMS Conference:
York 2000 — Modelling Biomolecular Mechanism***April 5-8, 2000**University of York, York UK*

This meeting focuses on computational approaches to investigating conformational and reactive processes in biomolecules, covering theory, methods, and applications. A key aspect of the meeting is to explore the interface with experimental approaches, highlighting the application, and prospects of single molecule and time-resolved techniques. Session topics include:

- Pathway Methods
- Enzyme Reactions
- Transport and Diffusion
- Folding
- Manipulation
- Energy Transfer and Transduction
- Solution Dynamics
- British Biophysical Society Invited Speaker

Contact:

Dr. Leo Caves
c/o Caroline Myers
Department of Chemistry
University of York

*York YO10 5DD UK**caves@ysbl.york.ac.uk**www.mgms.org/york2000/***RECOMB2000, The Fourth International Conference on Computational Molecular Biology***April 8-11, 2000**Tokyo Big Sight, Tokyo, Japan***Contact**

Human Genome Center
University of Tokyo
Tokyo, Japan
recomb2000@ims.u-tokyo.ac.jp
<http://recomb2000.ims.u-tokyo.ac.jp>

**Drug Discovery Technologies 2000:
Accelerating the Selection and Validation of Winning Compounds***April 10-12, 2000**Convention Centre Basel, Basel, Switzerland***Contact**

IBC UK Conferences
www.ibcuk.co.uk

Millennial World Congress of Pharmaceutical Sciences*April 16-20, 2000**San Francisco, CA USA*

The first World Congress of Pharmaceutical Sciences will be held at the San Francisco Convention Center, mid-April in the year 2000. The meeting will be coordinated by a 12-person organizing committee of scientists representing North America, Europe, and Japan. The meeting will be run under the auspices of the Board of Pharmaceutical Sciences (BPS) of the International Pharmaceutical Federation.

The meeting will include 42 half-day symposia, scheduled from Monday morning through Thursday noon, a number of update lectures, more than 1,200 anticipated poster presentations over the three days, an opening session, and a reception on Sunday.

Scientific symposia will include core topics of:

- Analytical and bioanalytical chemistry
- Drug metabolism/biochemical pharmacology
- Basic and clinical pharmacokinetics/dynamics
- Drug delivery/drug targeting
- Medicinal chemistry, natural products, and drug design
- Pharmaceutical biotechnology
- Formulation research/pharmaceutical technology
- Drug: membrane/receptor/transport interactions
- Bioavailability/bioequivalence

Contact*www.pharmweb.net/pwmirror/pwj/mill.html*

Royal Society of Chemistry Annual Congress

April 17-20, 2000
UMIST, Bridgewater Hall, Manchester, UK

The conference attempts to portray chemistry from the widest possible perspective in a way which is relevant to all RSC members, to promote collaboration and encourage new areas of research spanning the classical divisions of chemistry. A key theme of this millennial Annual Conference is to look to the future, whether this be exploring chemistry at the interface with biology, developing clean and sustainable processes to reduce the environmental impact of our industry, or the quest for nanoscale devices.

Contact

Nicole Morgan
2000 Annual Conference, Conferences Department
The Royal Society of Chemistry
Burlington House, Piccadilly
London W1V 0BN UK
Tel.: + 44 171 437 8656, Fax: + 44 171 734 1227
conferences@rsc.org
www.rsc.org/lap/confs/annconf2000.htm

May 2000

Second Indo-US Workshop on Mathematical Chemistry

May 30-June 3, 2000
University of Minnesota Duluth, Duluth, MN USA

The workshop will bring together leading researchers in the field of mathematical and computational chemistry. The results of latest research and applications of mathematical and computational chemistry in drug discovery, environmental toxicology, quantitative structure-activity relationships (QSAR), quantitative molecular similarity analysis (QMSA), chemoinformatics, and bioinformatics will be discussed. Dilip K. Sinha, Visva Bharati University (India) and Subhash C. Basak, Natural Resources Research Institute, University of Minnesota Duluth, are the co-chairs.

Contact

Subhash C. Basak, Workshop Co-Chairperson
Natural Resources Research Institute
University of Minnesota Duluth
5013 Miller Trunk Highway
Duluth, MN 55811 USA
Tel.: 218-720-4230, Fax: 218-720-4328
sbasak@wyle.nrri.umn.edu
<http://wyle.nrri.umn.edu/IndoUS2/>

June 2000

10th International Congress of Quantum Chemistry

June 5-10, 2000
Palais de l'Europe

8 avenue Boyer
Menton, 06500 France

Organizing committee

- Ernest R. Davidson
- Nicholas C. Handy
- Sigrid D. Peyerimhoff
- Alberte Pullman
- Jean-Louis Rivail
- Bjorn Roos

Contact

Xth ICQC 2000
Laboratoire de Chimie Theorique
Universite Henri Poincare-Nancy I B. P. 239
F- 54506 Vandoeuvre Les Nancy Cedex France
icqc@lctn.u-nancy.fr
www.lctn.u-nancy.fr/ICQC

July 2000

The Annual Meeting of the American Crystallographic Association

July 22-27, 2000
RiverCentre, 175 West Kellogg Blvd., St. Paul, MN USA
www.rivercentre.org

The ACA Transactions Symposium will be on using crystallography to understand biological mechanisms, encompassing molecules of all sizes. Two full-day workshops are planned for Saturday, July 22; SHELX for twins and macromolecular structures; QUEST; and one half-day workshop on Making Technical Presentations. Technical sessions will begin on Sunday, July 21, with hot new structures, refinement at ultra-high resolution, battery materials, and advances in small-angle scattering. Sessions continue through Thursday, with cool structures, general interest, new science from neutron sources, protein-nucleic acid interactions, high throughput crystallization, service crystallography at synchrotrons, crystal engineering (dedicated to Peggy Etter), science at long-length scale, nuclear industry materials, problem structure determination, and network glasses.

Contact

American Crystallographic Association
P.O. Box 96, Ellicott Station
Buffalo, NY 14205-0096 USA
Tel.: 716-856-9600 ext. 379, Fax: 716-852-4846
aca@hwi.buffalo.edu

August 2000

ISMB2000: 8th International Conference on Intelligent Systems for Molecular Biology

August 20-23, 2000
University of California San Diego, La Jolla, CA USA

This Conference provides a general forum for disseminating the latest developments in bioinformatics. ISMB is a

multidisciplinary conference that brings together scientists from computer science, molecular biology, mathematics and statistics. Its scope includes the development and application of advanced computational methods for biological problems.

ISMB 2000 will place special emphasis on knowledge discovery from the modeling and simulation of complex biological systems. This includes interpretation of large-scale gene expression data, whole genome comparative analysis, mathematical modeling of biochemical pathways, and interpretation of large macromolecular assemblies using data at different resolutions.

Relevant computational techniques include:

- Machine learning
- Pattern recognition
- Knowledge representation
- Databases, combinatorics
- Stochastic modeling
- String and graph algorithms
- Linguistic methods
- Robotics
- Constraint satisfaction
- Parallel computation

Biological areas of interest include:

- Molecular structure
- Genomics
- Molecular sequence analysis
- Evolution and phylogenetics
- Molecular interactions
- Metabolic pathways
- Regulatory networks
- Developmental control
- General molecular biology

Contact

Nancy Jensen

Tel.: 858-534-5039

ismb00@sdsc.edu

http://ismb00.sdsc.edu

September 2000

Structural Genomics: From Gene to Structure to Function

September 20-22, 2000

Robinson College, Cambridge, UK

Structural Genomics promises to be one of the most important subjects in the post-genome sequencing era. The availability of completed genomes, in addition to recent advances in method for three-dimensional structure determination mean that structure data will be available for more proteins

than ever before. As protein structure is central to almost all aspects of molecular biology and disease, Structural Genomics will have a great impact on the way we address biological problems.

This conference aims to discuss the major issues and developments related to Structural Genomics, including target selection, protein expression, functional genomics, structure determination, and bioinformatics.

In addition to talks from a range of distinguished speakers, the conference will also have poster and commercial exhibitions.

Preliminary Speaker List

(confirmed as of 10 November 1999)

- Sung-Hou Kim (Lawerence Livermore National Laboratory, CA USA)
- John Moult (CARB, Maryland, USA)
- Erik Sonnhammer (Karolinska Institute, Stockholm, Sweden)
- Guy Montelione (Rutgers University, NJ USA)
- Christian Cambillau (AFMB-CNRS, Marseille, France)
- Aled Edwards (University of Toronto, Canada)
- Hartmut Oschkinat (Research Institute For Molecular Biology, Berlin, Germany)
- Iain Campbell (University of Oxford, UK)
- Victor Lamzin (EMBL, Hamburg, Germany)
- Udo Heinemann (Max-Delbrück Centre, Berlin, Germany)
- Richard Durbin (Sanger Centre, UK)
- Mike Sternberg (Imperial Cancer Reseach Fund, London, UK)
- David Jones (University of Warwick, UK)
- Pete Artymiuk (University of Sheffield, UK)
- Sarah Teichmann (MRC, Cambridge, UK)
- Annabel Todd (University College, London, UK)
- Geoff Barton (European Bioinformatics Institute, UK)
- Nicolas Guex (GlaxoWellcome, Geneva, Switzerland)

Program Committee

- Guy Dodson (York University/Mill Hill, London, UK)
- Richard Durbin (Sanger Centre, Cambridge, UK)
- Sherin Abdel-Meguid (SmithKline Beecham, UK)
- Robert B. Russell (SmithKline Beecham, UK)
- Janet M. Thornton (University College, London, UK)
- Gabriele Varani (Laboratory of Molecular Biology, Cambridge, UK)

Organizing Committee

- Robert B. Russell (SmithKline Beecham, UK)
- Evelyn Boyle (SmithKline Beecham, UK)

Contact

www.mgms.org/cambs2000/index.htm

ACS COMP PRIZE

\$1,000 Prize to be Given for Best Talk at the ACS Computers in Chemistry Division's Emerging Technologies Symposium in Washington, DC

The COMP Division is organizing a special symposium of contributed papers at the Washington, DC, national meeting of the American Chemical Society, August 20-24, 2000. US\$1,000 will be given to the best oral paper presented at the symposium.

You are invited to participate! The symposium is open to everyone.

The papers will be evaluated based on the impact the research will have on the future of computational chemistry and allied sciences.

To participate, e-mail a 1,000-word abstract to the organizer by April 12, 2000. The abstract must be in text-only, MS Word, or RTF format. These long abstracts will be reviewed by a panel of experts. The best contributions will be selected for oral presentations at the meeting in Washington. The remaining contributions will be presented at a special evening poster session. It is essential that you also submit a short regular ACS abstract to the web-based system at <http://www.acs.org/meetings/> by April 12.

At the meeting, the oral presentations will be judged by the panel of experts, and the winner of the \$1,000 prize will be selected and awarded following the talks.

For more information, contact the organizer:

Prof. Donald B. Boyd
Organizer, Emerging Technologies Symposium
Department of Chemistry
Indiana University-Purdue University at Indianapolis
(IUPUI)
402 North Blackford Street
Indianapolis, IN 46202-3274, USA
boyd@chem.iupui.edu
Tel.: 317-274-6891
Fax: 317-274-4701