

## مدل‌سازی ملکولی و طراحی منطقی دارو

استانی سوانسون و کندراوریک

ادگار مایر<sup>\*\*</sup>

آنتونی<sup>\*</sup> لکزکویستی

چکیده:

پیشرفت‌های شایانی در مطالعه ماکرو ملکولهای زیستی به خاطر دو متده است، کریستالوگرافی به کمک اشعه  $X$  و گرافیک کامپیوترا حاصل شده است. ابزارهای لازم برای دسترسی به اطلاعات و بکارگیری این اطلاعات به کمک اینترنت و شبکه جهانی کامپیوترا در دسترس همه می‌باشند. در این مقاله کاربرد شبکه جهانی کامپیوترا توسط یک انسجو و مقایسه نتایج حاصل با ساختمان حاصل از تجربه دو متالوپروتئینی انسانی مورد بحث قرار خواهد گرفت. علاوه بر آن یک برنامه گرافیکی مدل‌سازی ملکولی *Pronto*، نیز شرح داده خواهد شد. در نهایت این مسئله مورد بحث قرار خواهد گرفت که چگونه پیشرفت علوم ممکن است قادر باشد مشکلات جاری در زمینه طراحی منطقی دارو را حل نماید.

۲- طراحی دارو  
۴- نگاره‌سازی رایانه‌ای

کلید واژه‌ها: ۱- مدل‌سازی ملکولی  
۳- بلورنگاری با اشعه  $X$

\*- بخش بیوفیزیک، انتستیتو فیزیک تجربی دانشگاه ورشو، لهستان

\*\*- آزمایشگاه بیوگرافیک، بخش بیوشیمی و بیوفیزیک دانشگاه تگزاس، تگزاس، امریکا

# MOLECULAR MODELING AND RATIONAL DRUG DESIGN - A PERSPECTIVE

A. Laczkowsky\*

Edgar F. Meyer\* #

Stanley M. Swanson#

R. Kendra Worick#

## ABSTRACT

Enormous progress had been made in the study of biological macromolecule primarily due to two essential methods, X-ray crystallography and interactive computer graphics. The tools necessary to access and utilize these data are at our fingertips thanks to the World Wide Web and Internet. We here illustrate the use of the web by a beginning student and compare the results with experimentally determined structures of two human matrix metalloproteinases. We then discuss a facile graphics modelling program, PRONTO. Finally, we speculate on what science might be like when some of the current bottle-necks are removed.

**key words:** 1) *Molecular Modeling*      2) *Drug Design*  
                  3) *X-ray Crystallography*      4) *Computer Graphics*

## INTRODUCTION

Somewhat over a century ago (1869), the "alphabet" of chemistry was codified by Mendeleeff (12). Since then, the vocabulary and grammar of chemistry have been developed to let us create virtually any molecule, natural or synthetic. Barely 100 years ago, peering into the darkness of the future, Emil Fischer postulated <sup>(7)</sup> that

molecules had specific 3-dimensional structures and that the structures could fit (or not) like "lock and key", the leitmotif of structure-based drug design. On the day of this lecture, 100 years ago (November 8, 1895), the light to illuminate this darkness, X-rays, was discovered by Wilhelm Röntgen (19). The resulting crystallographic method has been focused on virtually all major

\* Department of Biophysics, Institute of Experimental Physics, University of Warsaw, Zwirki i Wigury 93, PL-020-089 Warszawa, Poland

# Biographics Laboratory, Department of Biochemistry and Biophysics, Texas A & M University, College Station, Texas 77843-2128 USA

classes of biomolecules which can be crystallized and the "lock and key" simile has been amply confirmed. This possibility was first realized by the discovery of X-ray diffraction by von Laue (22) and this students in 1912, a discovery that completely changed our understanding of chemistry and biology. The language of the library of life (the double helix (23) and the genetic code) has been known for less than half a century. Taken as a book, this knowledge is best illustrated by computer graphics models of molecules; through it we continue to discover which volumes, which pages, and which passages of this library are relevant to specific vital processes. The read-out of the genome, as it proceeds, permits the identification of specific genes that can be linked to pathological conditions (e.g., human breast cancer, schizophrenia, obesity, etc.) While the knowledge derived from our increased understanding of biological processes brings responsibilities, and permits and even requires decisions not imagined a few years ago, it also forces us to come to terms with the ultimate eventuality that each individual could be subject to a complete readout: Preventive medicine in the future will start with a search of individual libraries.

Let us therefore review where we are now, how we can deal with current challenges, and from this vantage point try to peer into the future - what might science, medicine, or

pharmacology be like 100 years from now - what questions will then challenge our successors and what tools will they have at their disposal and how primitive will the elegance of our methods appear a century from now?

For several decades, it has been possible to determine the structures of specific molecules to high accuracy (by X-ray diffraction, by NMR spectroscopy, and, in the case of small molecules, by computational methods). Monochromatic (10) and color (15) molecular graphics makes it possible to visualize and manipulate models of molecules, thanks to ready access to the Protein Data Bank (1). While models of spherical atoms and rigid bonds represent a time-averaged molecular model, molecular dynamics methods (with experimental constraints) permit (11) the investigator to probe internal flexibilities and thus consider a greater range of alternative scaffolds and binding modes. The breadth and depth of these images of the molecules of life, and our strained ability to perceive and manipulate them, arch past Fischer (7) and Lucretius (21) into the depths of Plato's cave (18) and the colorful shadows that catch our fancy. While the genetic code allows us to relate sequences of nucleic acids to sequences of proteins, the second code, protein folding, has yet to be broken. When it is, and this is only a matter of time, the volumes in a person's library can be envisioned as

functioning (or malfunctioning) molecules.

Within families of molecules, it is now possible to utilize predictive methods to identify conserved domains of sequences and relate them to known structures. So, let us put this in the context of a beginning student with a textbook knowledge of biochemistry. This student could be in Cambridge or in Cairo, (in Tulsa, Tehran, or Thule); the Internet linkage is the same and thus provides access to the same libraries and the same tools to link information to knowledge to experiment.

Over the past few weeks, a beginning biochemistry student (R.K.W.), besides attending lectures, discovered for herself how to use the tools of Internet to extract the sequence<sup>(5)</sup> of mouse MMP-13 and predict its (unknown) structure. In collaboration with Professor Yves Eeckhout (Université Catholique de Louvain Faculté de médecine), we are in the process of crystallizing this enzyme and will subsequently determine its structure. The prediction method already has been validated<sup>(16)</sup> and is generally applicable when the preconditions are met. Moreover, as this student has demonstrated, it does not presuppose a residency at one of the dominant scientific centers. But such knowledge can be overwhelming. Already the 4000 protein structures in the Protein Data Bank<sup>(1)</sup> exceed the perceptual capacity of a single individual. The average "small" (200-250 amino acid) protein contains so

much information that answers greatly exceed questions; but satisfaction can be derived from observing additional relationships, as simple as "obvious" structure - function relationships of internal water molecules and linked H-bonds<sup>(13)</sup> or as complex as unexpected and backwards binding<sup>(14)</sup>. This strongly suggests that other relationships await discovery and application to one of the exciting challenges before us: to understand physiological function at the molecular (atomic, electronic) level and to use available knowledge of structure-function relationships to design highly specific and potent compounds to treat molecular diseases.

While computer literacy and, for the most part, a functional knowledge of English are currently a prerequisite, the student in Texas, or in Tehran, has equal access to these data bases and the tools to use them. So let us consider a current research topic and the results these tools provide.

### Prediction of the structure of a matrix metalloprotein

Collagenases are proteolytic enzymes capable of cleaving specific peptide bonds in triple helical collagen. Uncontrolled proteolysis by collagenases has been linked to pathologies such as arthritis, cancer, periodontitis. Collagenases has been linked to pathologies such as arthritis, cancer,

periodontitis. Collagenase inhibitors which bind to the active site Zn atom have been shown to control penetration of the extracellular matrix without killing the cancer cell or interfering with the genetic expression in terms of m-RNA levels<sup>(6)</sup>. The truncated, 160 amino acid metalloenzyme remains the minimal pharmaceutical target, and several industrial and academic laboratories are actively pursuing the design, synthesis, and testing of novel compounds which inhibit this class of enzymes in the mM to nM range. Besides the published structures of the mammalian collagenases identified by the signature sequence, HExxHxxGxxH, snake venom and bacterial collagenases have been found to possess a similar active site scaffold. Another common feature, unique among the proteinases, is the conserved S1<sup>(12)</sup> primary specificity site<sup>(8)</sup> which is much larger than the side chains of any of the 20 amino acids. For example, the indole ring sidechain of *Trp*<sup>(24)</sup> leaves this pocket half-empty (otherwise filled with water), so no peptide sidechain can fill this pocket completely. Yet, nature had gone to great lengths to conserve this cavernous cavity among bacterial, snake venom, and mammalian enzymes. Why? We don't know. We postulate<sup>(4)</sup> that some endogenous inhibitor has yet to be

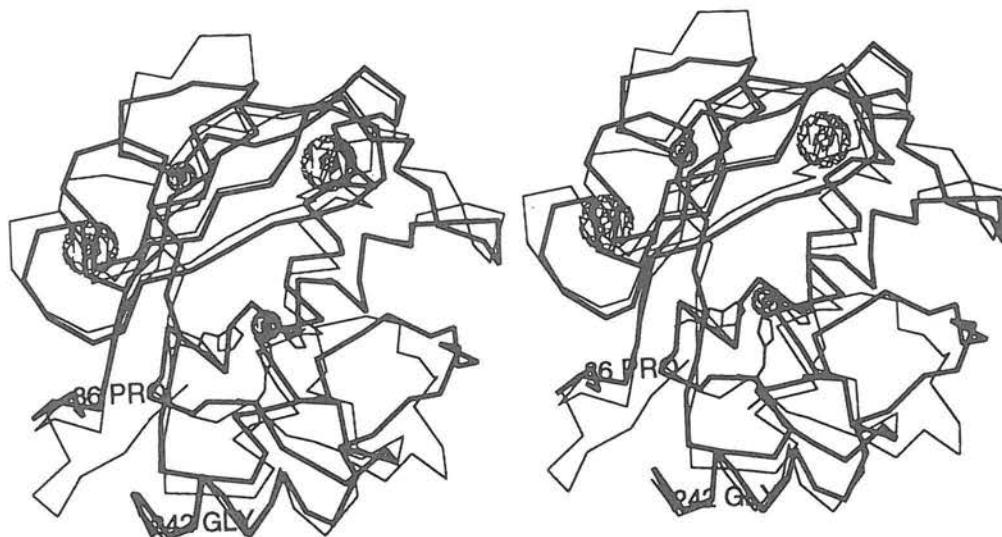
discovered, if we are serious in assuming that nature has a purpose linked to a property. This may be the source of yet another surprise, which would keep the study of these molecules from becoming routine.

Using program SWISS-MODEL<sup>(17)</sup> available on the WorldWideWeb or Internet (<http://expasy.hcuge.ch/swissmod/> SWISS-MODEL.html), one of us (R.K.W.) brought the sequences of the structurally known and unknown collagenases into registry\* with human MMP-8<sup>(2)</sup> and predicted the structure of the unknown enzyme, mouse MPP-13. The refinement option of program SWISS-MODEL was then utilized to cause backbone bond lengths and angles to conform to standard values. The resulting structure was returned via Internet for inspection and evaluation. Using the whole-molecule translation-rotation function (FBRT) of program PRONTO, the model of MMP-13 was superimposed visually with the experimentally determined structure of MMP-8<sup>(2)</sup>. The similarity was striking, all the more because the experimentally determined target structure contained two Ca and two Zn atoms, one of which is the catalytic Zn in the active site. The prediction method had no knowledge of Ca or Zn ligation, yet the folding of the backbone in the ligated regions does not

\* This was the tricky part; a single shift, out of registry by one amino acid, prevented the program from making a meaningful prediction.

differ dramatically from other regions. While the visual superposition procedure is somewhat subjective, a least-squares superposition procedure, RIGID (R.J. Remington, unpublished) in PRONTO gave a root-mean square (RMS) superposition\* agreement of  $0.56 \text{ \AA}$ . Visual inspection of

Fig. 1 confirms this structural homology. To test the extent of this homology, the same sequence of MMP-13 was likewise superimposed with that of fibroblast collagenase, MMP-1<sup>(3)</sup>, with a RMS fit of  $0.61 \text{ \AA}$  (Fig. 2)\*\*. The backbone ( $\text{C}\alpha$ ) positions of all three structures are further



*Figure 1. A stereo view prepared with program PRONTO depicting the  $\text{C}\alpha$  backbone structure of the target enzyme, MMP-8 neutrophil collagenase (heavy lines, from 86 to 242), over which was superimposed (RIGID) the  $\text{C}\alpha$  structure of the predicted model of mouse MMP-13 (thin lines). The Zn and Ca atoms of human MMP-8 are depicted as small and large spheres the active site His residues are drawn as thin lines near the Catalytic Zn atom. The N terminus (Pro 86) and the carboxyl terminus (Gly 242) are labelled. One notices that the MMP-13 N-terminus points towards the active site, which extends from left to right in these views.*

\* Because of the conserved HExxHxxGxxH sequence, which comprises the active site helix-bend-His structure, the  $\text{ca}$  atomic positions of those plus four additional helical amino acids were superimposed; no other structural constraints were applied.

\*\* The largest differences are found at the penultimate  $\text{Ca}$  positions ( $1.41 \text{ \AA}$  and  $1.66 \text{ \AA}$ , respectively); if these two  $\text{Ca}$  positions are omitted, the respective RMS values are  $0.42 \text{ \AA}$  and  $0.43 \text{ \AA}$ , respectively.

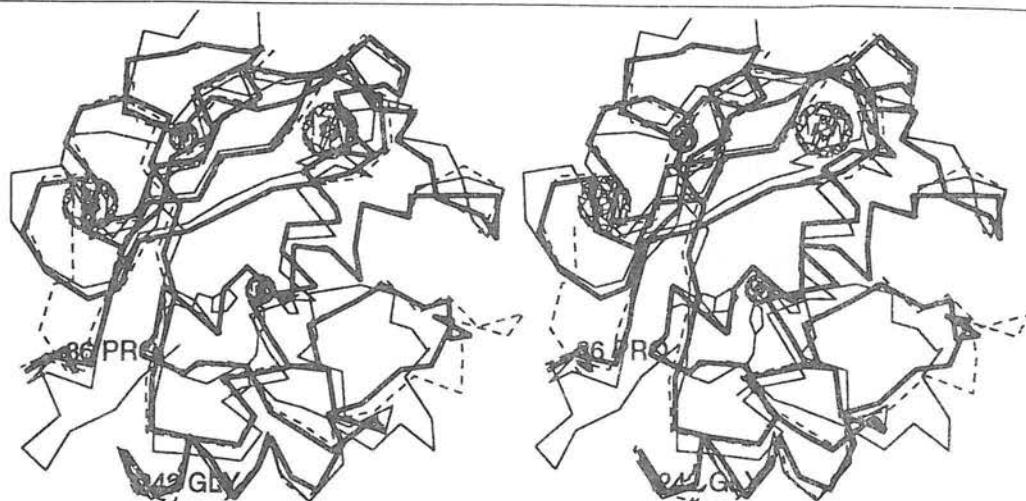


Figure 2. A stereo view of the experimentally determined model of human fibroblast collagenase, MMP-1 (dashed lines), superimposed with the RIGID command over a  $C\alpha$  model of human MMP-8.

superimposed in Fig. 3. these results confirm the current capabilities of Internet to link data bases and extract knowledge from information; they also demonstrate how a

beginning student can make use of these tools to obtain a unique prediction. This could have been done anywhere in the world with a link to Internet!

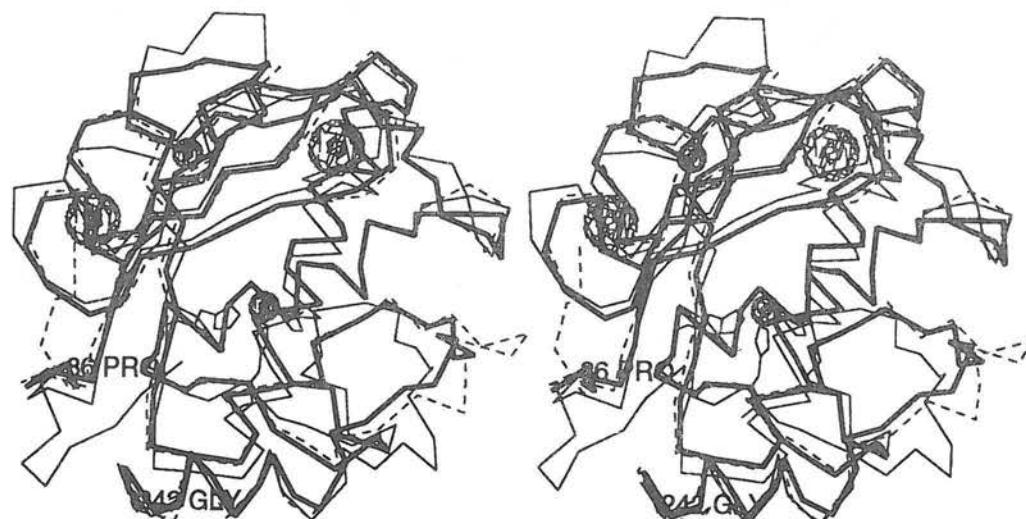


Figure 3. A stereo trace of three  $C\alpha$  structures is superimposed with the aid of the RIGID routine. The target enzyme, human MMP-8, is drawn as heavy lines; the  $C\alpha$  trace of human MMP-1 is drawn as dashed lines; both are experimentally determined structures. The predicted model of mouse MMP-13 is drawn as thin lines. The only constraint was the least squares fit of the active site  $C\alpha$  positions of the three structures. The catalytic Zn and structural Zn and Ca atoms are drawn as spheres. One can see that the general topological similarity is striking, the greatest differences occurring at several external loops.

## PRONTO

Historically, interactive computer graphics modelling tools came from an academic environment in the 1960s (i.e., at MIT, Princeton, Washington University, Texas A&M), sites. Program FRODO (9), developed at the Max-Planck-Institut für biochemie, was the first all-purpose modelling program, both for its extensive functionality but also because it was implemented using readily available commercial hardware. The program grew and matured, thanks also to input and modifications from a number of users, into the most widely used program for the determination and analysis of macromolecular structures. By challenging this and other programs to utilize new features in hardware and software, several advances of hardware, most recently the graphics workstation, have made older graphics systems like FRODO obsolescent, in spite of their functionality and utility. The opportunity for "open systems" development of software has been met with a spotty response in the industry, further complicating the portability of programs. Most recently, the pronounced trend towards the dominance of commercial modelling packages has indeed made computer graphics modelling widely available as well documented turn-key system, but at a significant price in terms of stifling the development and implementa

tion of novel features: The vendors do not release the source code.

*PRONTO* is a port of FRODO (A. Jones) and PSFRODO (F. Quiocco) to the Unix graphics workstation environment. It was conceived by E. Meyer and W. Steigemann as a "one-time" port of the graphics interface to an industry standard open system (PEX), an extension of X windows three dimensional graphics (PHIGS+), so that transferal of the code to new workstations would not be as labor intensive as previous ports of FRODO had been. It thus preserves a familiar and extensively developed modelling tool idea was to make source code available to the research community so that algorithms could be understood, corrected, or extended locally. Some thought was given to file compatibility between platforms, which remains a consideration with the cumbersome use of similar but different formats (PROTEIN, CCP4, etc.)

The initial design and coding was performed (A.L.) on an Evans and Sutherland ESV system. Intensive debugging (S.M.S.) at MPIB-Martinsried and subsequently here in Texas resulted in a usable crystallographic modelling system which has also been used by several sets of beginners in a graduate class in structural biochemistry. Ports to a Silicon Graphics workstation were recently undertaken. Three different 3-D environments were partially or entirely implemented: PEXlib,

figaro+ (a proprietary multiplatform PHIGS+), and native GL. Two to 4 weeks of programming time per environment were required to obtain useable but not completely debugged code. Because this was the first port to a different workstation, extra time had to be spent in identifying and segregating machine and graphics specific parts of the code. PRONTO retains PHIGS calls for graphics, but these are translated to the other 3-D graphics protocols in a set of interface routines. Some of the graphics objects had to be changed to make interfacing workable and efficient. Difficulties appeared where industry standardization has not occurred: coordination of window definition between X and the 3-D context, implementation of stereo, color definitions, and in the interactive areas of picking and dials. Subsequent ports should not be as difficult. Current efforts are directed to maintenance debugging and to incorporation of core-tracing for density skeletonization<sup>(20)</sup>.

The current implementation of PRONTO contains a basic 3-D comparator for a model, 8 density maps and 16 static objects (MOL files). Model manipulation is by 6 torsion angles, rigid fragment movement (rotation and translation: FBRT), and single atom movement. Model atoms can be labelled and their geometry determined. Subimages are switchable by menu buttons or by "hot keys". Control of color and

intensity of subimages is interactive. Side-by-side and hardware stereo imaging is used routinely, in addition to monocular viewing.

While the experimental basis for modelling is clearly understood among the chemical community (including some of its complexities,<sup>15</sup>) little is said or done about the computational and graphical side. The combination of structure prediction and molecular comparison is a straightforward example of the current capability of an "open-system" approach. The mix and match of functions is left to the creative powers of the individual and the graphics program remains open-ended in its implementation.

Where information is freely available, knowledge will flourish. Commercialization of this process fundamentally threatens academic process of teaching and research. An incremental charge for access to the Internet files like Swissprot or the Protein Data Bank, or charge for a successful protein structure prediction by SWISS-MODEL would have created barriers prohibiting the exploratory research described above. With charges for commercial molecular graphics systems comparable in magnitude with a graduate student stipend, and with funds for teaching and research becoming increasingly harder to obtain a major re-evaluation of priorities may be necessary in order for academic

research to continue to prosper. Commercialization of software has not helped advance the cause of academic science.

## PREDICTIONS

Imagine a 3-D graphics display the size of a notebook. Imagine being able to talk to it and it to you, in your language. Imagine being able to take it from room to room, on a train or plane, and use it without connecting wires. Imagine a high-speed linkage to the world. Imagine a neural linkage, so that thoughts could be perceived and integrated - could our brain tolerate the competition - and use it to stimulate our creativity? or replace it?? or us?? Imagine a neurological (and electronic) linkage of 2 or more brains (massively parallel processing = the neural committee). Imagine then being able to repair or regenerate severed nerves or stimulate the immune system to treat a viral infection. Imagine an atomic-resolution map of the surface of a functioning cell, organelle, ribosome, muscle fiber, organ. Imagine a chemical or computational link to learning and memory, rendering the classroom obsolete. If these advances could detect and cure human illness, might they also remove ignorance-based hatred or frustration and help us progress in the search for human destiny. If we were capable of achieving even some of these objectives, would we be worthy of them? Could our moral and value systems expand

and encompass cohabitation with a sleepless computer? can anyone with a digital watch doubt that we have already? It is only a matter of degree.

The visual component of interactive computer graphics reaffirms that our methods are still at a relatively primitive level, barely beyond the capabilities of the standard television monitor. A principal problem is dimensionality; the 2-dimensional surface of the CRT display creates successive 2-D images, whereas our visual system is designed to function in a 3-D world. Several tricks indeed make it possible to bridge this information and dimension gap, but they are not widely available. Until high band-width systems with neural linkages make graphics hardware obsolete, an holographic display is needed to create realistic 3-D images. The folding problem may be solved before the visualization problem.

## CONCLUSION

After reviewing the historical basis of molecular pharmacology, we have shown how a beginner could use available computational tools to create a realistic model of a novel macromolecule. We then discussed the need for facile graphical tools to visualize and manipulate complex molecular models. Finally, on a flight of fancy, we sought to project these ideas around the world and into the future. Others

are encouraged further to facilitate the free flow to fruition of these ideas.

### ACKNOWLEDGEMENTS

One of us (E.F.M.) wishes to acknowledge

a travel grant from GlaxoWellcome; local expenses were provided (S.M.S.) by the Department of Biophysics, University of Warsaw, and the Max-Planck-Institut für Biochemie, Martinsried.

### REFERECE

- 1) Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi T. & Tasumi, M. (1977). "The Protein Data Bank" *J. Mol. Biol.*, 112, 535-542.
- 2) Bode, W., Relinemer, P., Huber, R., Kleine, T., Schnierer, S. & Tschesche, H. "The X-ray crystal structure of the catalytic domain of human neutrophil collagenase inhibited by a substrate analogue reveals the essentials for catalysis and specificity (1994) *EMBO J.* 13:1263-1269.
- 3) Borkakoti, N., Winkler, F.K., Williams, D.H., D'Arcy, A., Broadhurst, M.J., Brown, P.A., Johnson, W.H. & Murray, E.J. "Structure of the catalytic domain of human fibroblast collagenase complexed with an inhibitor", *Nature Structural Biology* (1994) 1:106-110.
- 4) Botos, I., Zhang, D., Scapozza, L., Liotta, L.A. & Meyer, E.F., "Batinostat, A Potent Matrix Metalloproteinase Inhibitor with an Unexpected Mode of Binding", *Proc.Natl. Acad.Sci. USA*, in press.
- 5) Delaisse, J.M., Eeckhout, Y., Neff, L., Francois-Gillet, Ch., Henriet, P., Su, Y., Vaes, G. & Baron, R., "(Pro) collagenase (MMP-1) is present in rodent osteoclasts and in the underlying bone resorbing compartment". (1993) *J. Cell Sci.* 106:1071-1082.
- 6) Fife, R.S., Sledge, G.W. Jr. Effects of doxycycline on in vitro growth, migration, and gelatinase activity of breast carcinoma cells. *J Lab Clin Med* 125:407-411, 1995
- 7) Fischer, E. (1894). Einfluss der configuration auf die Wirkung der Enzyme. *Ber Deutscher Chem. Gesellschaft* 27, 2984-2993; n.b.p. 2992.
- 8) In the notation of I.Schechter & A. Berger, receptor subsites are denoted sequentially by S, juxtaposed peptide residues are denoted by P, counting both directions from the scissile bond, \*:  
---S3-S2-S1-S1<sup>1</sup> ----- +NH-P3-P2-P1\*  
P1<sup>1</sup>-P2<sup>1</sup>-COO-  
Schechter, I. & Berger, A. (1994) *Biochem. Biophys Res. Comm.* 27,157-162.
- 9) Jones, T.A. (1978) A graphics model building and refinement system for macromolecules *J. Appl. Cryst.* 11, 268-272
- 10) Levinthal, C., "Molecular

Model-Building by Computer" Scientific American 214:42-52 (June, 1966).

11) McCammon, J.A. & Harvey, S.C., "Dynamics of proteins and nucleic acids", Cambridge Univ. Press, Cambridge, 1987.

12) Mendeleff, D.I. (1905) "The Principles of Chemistry", Third English Edition, Longmans, Green, and Co., 39 Paternoster Row, London

13) Meyer, E. "Internal Water Molecules and H-bonding in biological Macromolecules: A Review of Structural Features with Functional Implications".(1992). Protein Science 1:1543-1562

14) Meyer, E.F., Botos, I., Scapozza, L.& Zhang, D., "Backwards Binding and Other Structural Surprises ", (1995), Perspectives in Drug Discovery and Design, 3:168-195.

15) Meyer, E.F. (1968) Annual Report, Brookhaven National Laboratory, p.8

16) Peitsch, M.C. & Jongeneel, C.V., (1993) "A 3-D model for the CD40 ligand predicts that it is a compact trimer similar to the tumor necrosis factors", Int. Immun. 5:233-238.

17) Peitsch, M.C. Bio/Technology (1995) 13:658-660

18) Plato, "The Republic" , book VII

19) RÖntgen, W.C. (1895) über eine neue Art bon Strahlen ", Sitz. Würzburger Phys-mediz. Ges., 3-12.

20) Swanson, S.M., "Core Tracing: depicting connections between features in electron density ", Acta Cryst. D50 695-708.

21) Titus Lucretius Carlus, " On the Nature of Things - De rerum natura " Book VI, in " The Stoic and Epicurean Philosophers ", Edited by W.J. Oates, Translated by H.A.J. Munro, Random House, New York, 1940.

22) Von Laue, M. (1912), Sitz. math. Phys. Klasse bayer. Akad. Wiss. p.303

23) Watson J.D. & Crick, F.H.C. (1953), "Molecular sturcture of nucleic acids ", Nature 171 737-8

24) Zhang, D., Botos, I., Gomis-Rüth, F.-X., Doll, R., Blood, C., Fox, J.W., Bode, W. & Meyer E.F., "Structural Interaction of Natural and Synthetic Inhibitors with the Matrix Metalloproteinase, Atrolysin C (Ht-d)". (1994) Proc. Natl. Acad. Sci. USA, 91:8447-8451