

Cavity search: An algorithm for the isolation and display of cavity-like binding regions

Chris M.W. Ho and Garland R. Marshall*

Center for Molecular Design, Washington University, St. Louis, MO 63130, U.S.A.

Received 19 July 1990

Accepted 26 September 1990

Key words: Cavity search; Ligand–receptor interaction; Molecular graphics; Receptor mapping; Solid modeling

SUMMARY

A set of algorithms designed to enhance the display of protein binding cavities is presented. These algorithms, collectively entitled CAVITY SEARCH, allow the user to isolate and fully define the extent of a particular cavity. Solid modeling techniques are employed to produce a detailed cast of the active site region, which can then be color-coded to show both electrostatic and steric interactions between the protein cavity and a bound ligand.

INTRODUCTION

Most substrate–enzyme or receptor–ligand interactions occur within pockets or cavities buried within proteins. Inside these cavities, a microenvironment is established that chemically favors the interaction between molecular species. Knowledge of the extent of such cavities can assist the study of binding interactions and the design of other binding ligands. The use of molecular modeling to display these interactions can offer insight as to the mechanisms behind these phenomena.

The visualization of substrate–enzyme or ligand–receptor interactions presents a formidable challenge. One must manipulate complex molecular structures in an attempt to understand their steric and electrostatic complementarity. However, it is difficult to discern spatial relationships merely from stick models, and this obviates an effective interpretation of the display [1]. The use of depth cueing and stereo to view molecular structures can improve the situation, but accurately judging the size and shape of the molecule remains difficult.

* To whom correspondence should be addressed.

Perhaps the innovation that has most greatly enhanced the visual study of protein binding cavities has been the development of algorithms to calculate and display molecular surfaces [2]. Such displays more realistically represent the volume that the structures occupy and provide immediately comprehensible information concerning steric complementarity. The well-known Connolly method [3] employs Richards' definition of a molecular surface [4] by rolling a probe sphere of given size over a molecule of interest, thereby generating a representation of the solvent-accessible surface. Other groups have augmented this method to display additional information. Barry introduced a technique where molecular surfaces of binding sites are calculated by adding an extra van der Waals radius to the normal surface (C.D. Barry, unpublished results). This allows one to collapse the surface of a receptor site onto a stick model of its ligand and eliminates the need to calculate the ligand surface. In doing so, one can approximate van der Waals contacts by observing regions where the ligand stick figure just touches the receptor surface. This simplifies the display while conveying the essential information. Weiner et al. color-coded molecular surfaces by electrostatic potentials calculated from quantum-mechanically derived partial atomic charges [5, 6]. Still others have color-coded molecular surfaces by hydrophobicity [7, 8].

There are numerous examples where these computer-generated displays have been employed to study the binding of entities within cavity-like structures [2]. In these studies, molecular surfaces were used to define the three-dimensional structure of the active site region. However, these regions are relatively small in comparison to the rest of the protein, and are usually formed by groups of residues located in different parts of the linear amino acid sequence [9]. As a result, a molecular surface display can become cluttered and distracting as extraneous atoms and bonds that support but do not form the active site remain in view. Even with clipping mechanisms to limit the depth of vision, one is still faced with uninteresting components of the protein appearing within the plane of section. Ideally, only those molecular surfaces in direct contact with the active site along with the volume that they contain should be displayed. This would produce a *cast* of the binding cavity. In this paper, we describe an algorithm to produce such a cast of the active site cavity, allowing the display of its internal volume without unnecessary structures.

Other groups have developed algorithms to find and display cavity-like regions of proteins. Kuntz et al. [10] developed a procedure to explore the steric complementarity between ligands and receptors of known structure. Using the molecular surface of a receptor, a volumetric representation of the binding cavity was approximated using a set of spheres that were mathematically 'packed' within it. Voorintholt et al. [11] used three-dimensional lattices to calculate density maps of proteins. In these maps, lattice points were defined as a function of the distance to the nearest atom. This technique is effective in delineating regions of low density where channels and cavities exist. However, problems arise in isolating and defining the extent of a specific cavity of interest. In any protein, there are numerous cracks, crevices, or gaps where spaces exist between atoms. In a three-dimensional density map, the cavity of interest can become obscured by these regions of low density. To circumvent this lack of specificity, our algorithm implements a search function to allow the investigator to isolate a single cavity of interest by specifying a seed point. From this seed point, the algorithm systematically explores the entire volume of the cavity, following its borders and effectively filling every crevice within it.

Once the cavity of interest is isolated, a *three-dimensional cast* of the internal volume is produced using techniques of solid modeling. A *solid model* is a mathematical description of a geometric object that is both complete and unambiguous [12–14]. Once defined as a solid model by

means of a solid modeler, an object may be rigorously manipulated as if it were physically real. The ability of solid modelers to perform Boolean operations upon defined models has been used extensively in studies of receptor mapping and host-guest docking [15, 16]. These techniques are invaluable tools in molecular modeling software, as pseudoelectron densities of atomic structures may be defined as solid models, manipulated, and then displayed using powerful imaging hardware and software.

The cast that is produced affords an uncluttered, yet highly detailed reproduction of the original cavity and allows an uncompromised view from any perspective. This display may then be manipulated interactively as if it were a molecular structure to analyze the binding of ligands. Electrostatic interactions as well as information regarding unoccupied regions of free space within the cavity may also be calculated and displayed. This will be elaborated upon in the discussion of the algorithms.

DISCUSSION OF ALGORITHMS

The components of a generic protein cavity must first be described. In Fig. 1, a three-dimensional cavity is depicted. Such a cavity may be described as being cup-like in shape, consisting of a *cavity wall* (in black) and enclosing a finite volume of space, *innerspace* (in grey). This innerspace volume is of interest, for it is here that various ligands bind and interact with the protein cavity. Innerspace is also a continuous entity. Any probe atom of a given size placed within one region should be able to traverse and access any other. Consider now the cavity wall itself. This wall may be of variable thickness; however, the fundamental principle is that the wall is an impenetrable

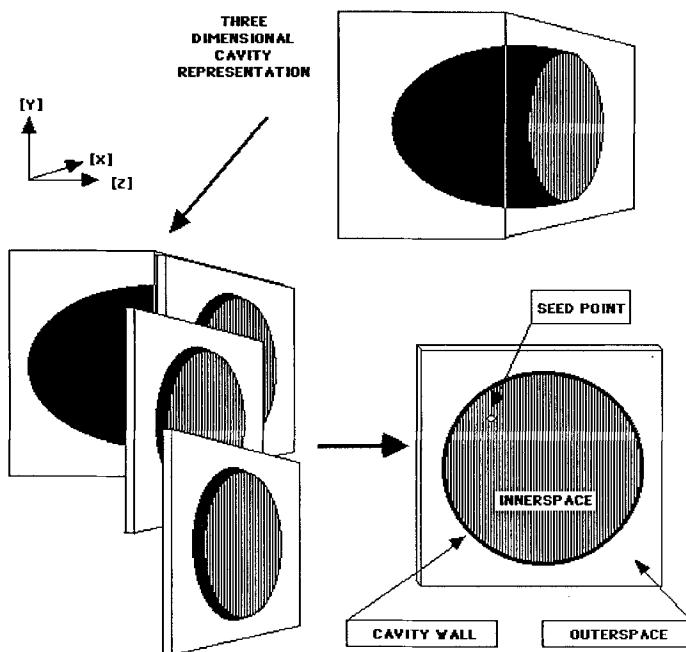


Fig. 1. A three-dimensional representation of a protein cavity.

entity, effectively containing the innerspace and separating it from the remaining regions, the *outerspace*.

The goal is to search and map out the three-dimensional shape of our specific cavity of interest. We also wish to isolate it from the myriad of cracks and crevices present throughout the rest of the protein. Consider the partitioning of the three-dimensional cavity into a set of serial sections, each of minute thickness. As depicted in Fig. 1, cross-sections of the cavity are obtained. In these cross-sections, elements of each of the three zones depicted above can be found: a region of innerspace depicted in grey completely bounded by a black cavity wall surrounded by the *outerspace* field. Any cross-section may be considered to be a *mold* of the innerspace region present at that corresponding location within the cavity. From a given *seed point* located within the innerspace area, a hypothetical filler substance poured into this mold and allowed to completely fill the innerspace and harden would maintain the shape of this region bounded by the cavity wall, thereby producing a *cast* of this particular portion of the cavity. Continuing this process for each serial slice in turn, an authentic reproduction of the cavity volume may be constructed by the alignment and reassembly of the individual casts as shown in Fig. 2.

The formation of a cast requires two components: a mold and filler. Physically, a cast is produced by introducing a filler substance into the mold and allowing it to harden. Upon its removal, the shape of the mold is preserved. Figure 3 shows how this process is reproduced mathematically using solid modeling techniques. The mold simply consists of the atoms comprising the cavity wall

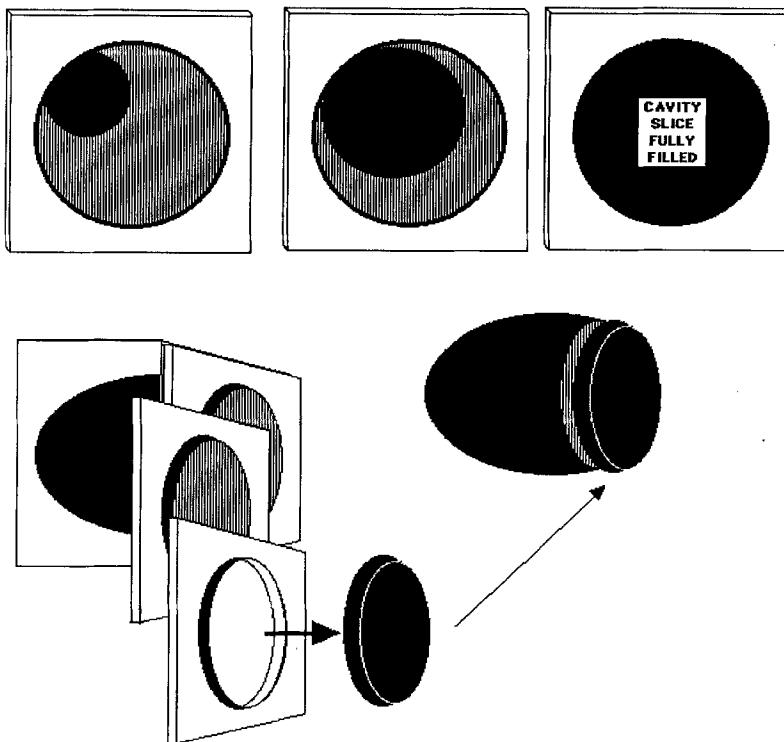


Fig. 2. Reconstruction of original cavity by realignment and joining of casts made from individual cross-sections.

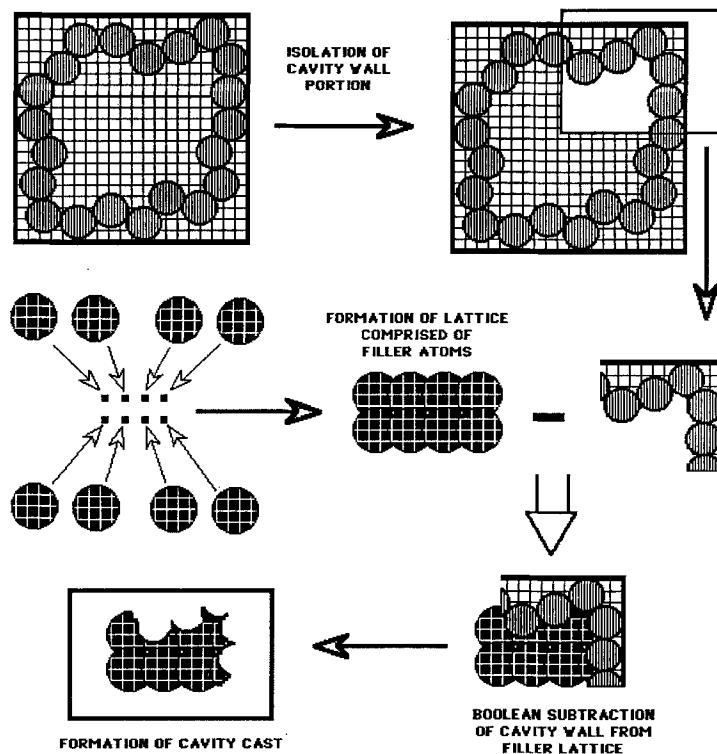


Fig. 3. The creation of molecular casts by solid modeling.

at that particular location. They are modeled as spheres whose relative sizes match the van der Waals radii of the atoms they represent. A solid model of filler is created using a *lattice* of atoms spaced close enough together that the union of their volumes produces a solid entity large enough to span the internal volume of the mold. This lattice approach to producing a large filler solid is used for two reasons. First, the SYBYL 5.3 molecular modeling system was used to perform these calculations [Tripos Associates Inc., St. Louis, MO, U.S.A.]. SYBYL 5.3 contains a solid modeling interface that allows the user to define solids using only *spheres*. This is not unexpected since the most common object shape one encounters in molecular modeling is the sphere. Second, the use of the lattice allows one to strictly define the extent of the filler solid. In Fig. 3, a portion of the cavity wall is isolated to demonstrate the casting process. It is imperative that the volume of the filler solid be made *slightly larger* than that of the cavity space. This slight overlap in volume ensures that the cavity mold is completely filled. By subtracting the volume of the cavity wall atoms from the solid filler volume, an exact cast of the shape of the cavity is produced.

There are two main functions that the filler solid must perform in any given slice. It must first spread out from the seed point and fully expand to fill every crack and crevice of the volume bounded by the cavity wall. Second, to ensure the formation of an exact cast, it must be slightly larger than the cavity volume and overlap with the cavity wall. To achieve this, we begin by superimposing a grid, the *search grid*, on a cavity slice which must be filled (Fig. 4). A standard two-di-

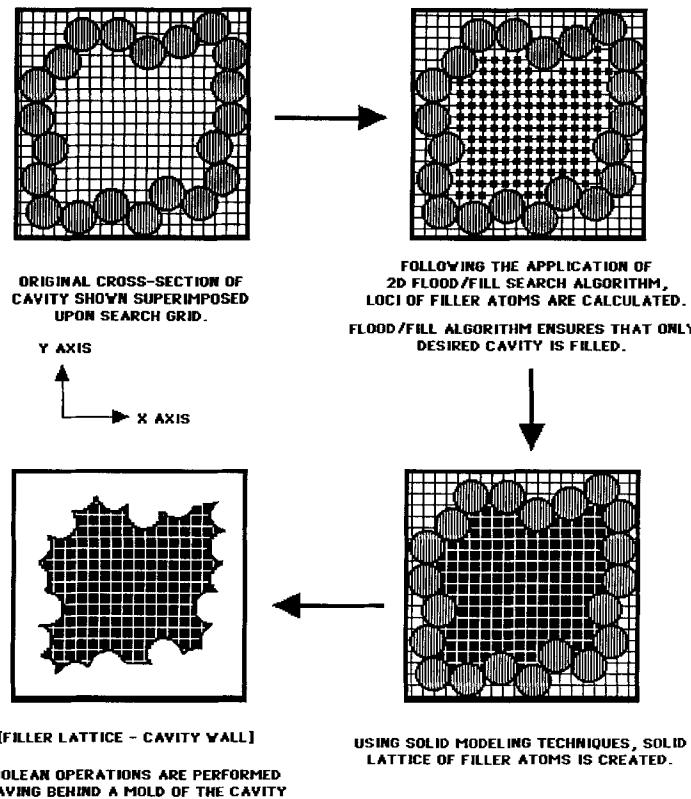


Fig. 4. The use of a flood/fill algorithm to completely fill and cast a single cavity cross-section.

dimensional *flood/fill algorithm* is then employed to search for and designate all *contiguous* grid points that lie within the cavity wall border. This algorithm is well documented [17] and is commonly used in drawing and painting software where objects are filled with a specific color or pattern. The algorithm requires the position of a *seed point*, which is the location from which the filler solid will spread out to fill any given slice. In our mold casting analogy, the seed point is akin to the location at which our filler is ‘poured’ into the mold, and it must be specified by the user as it designates which region is the cavity interior. As shown in Fig. 4, following the application of the flood/fill algorithm, a lattice has been formed comprised of all contiguous grid points lying within the boundaries of the cavity wall. These lattice points become the loci of filler atoms. The fineness of the search grid may be altered to suit the complexity of the cavity interior. A finer search grid will generally favor a more thorough search; however, more CPU time and memory will be required as a greater number of grid points will have to be analyzed and stored. To insure that overlap exists between the filler solid and the cavity wall atoms, it is imperative that the distance between grid points be less than the radius of the filler spheres as shown in Fig. 5. The filler solid is then created through the union of all the volumes of the spheres on the lattice. Subtracting the cavity wall from the filler solid produces a perfect cast of the cavity interior within this particular slice.

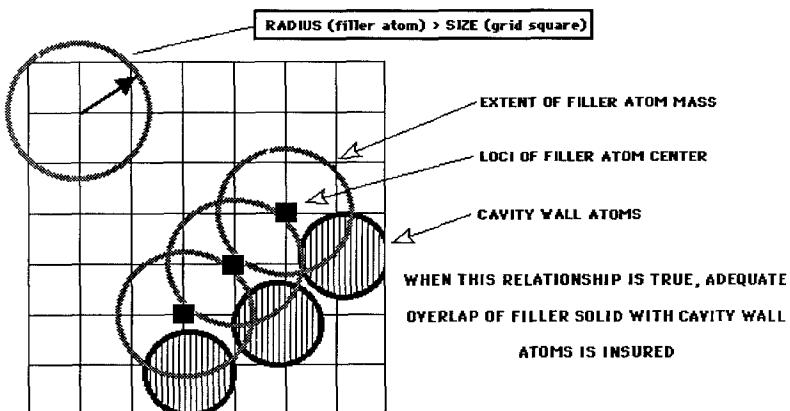


Fig. 5. The importance of maintaining a small enough grid size to insure that adequate overlap of the solid filler occurs.

The two-dimensional cavity slice filling process is detailed above and must now be expanded to the third dimension. Figure 6 shows how this is accomplished. For each lattice point found in slice N, the corresponding lattice points in the slices both directly above and directly below are checked to see if they fall in a 'clear' region. A clear region is an area that has not been searched previously and is not occupied by a cavity wall. If such a point is found, it can then serve as a seed point to begin filling this new cavity slice. This process is repeated until no new seed points are found. At this point, the entire internal volume of the cavity will have been explored.

Protein cavities contain gaps where the solvent, substrate, or ligand must enter. Figure 7 depicts such a situation. The orientation of the cavity is such that a cross-section contains an incomplete cavity wall. From a given seed point within the cavity, the flood/fill algorithm would normally in-

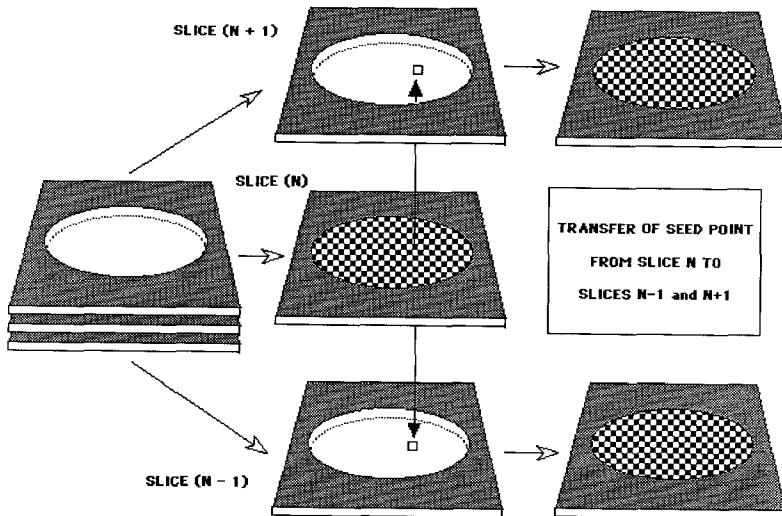


Fig. 6. Applying the flood/fill search to the third dimension.

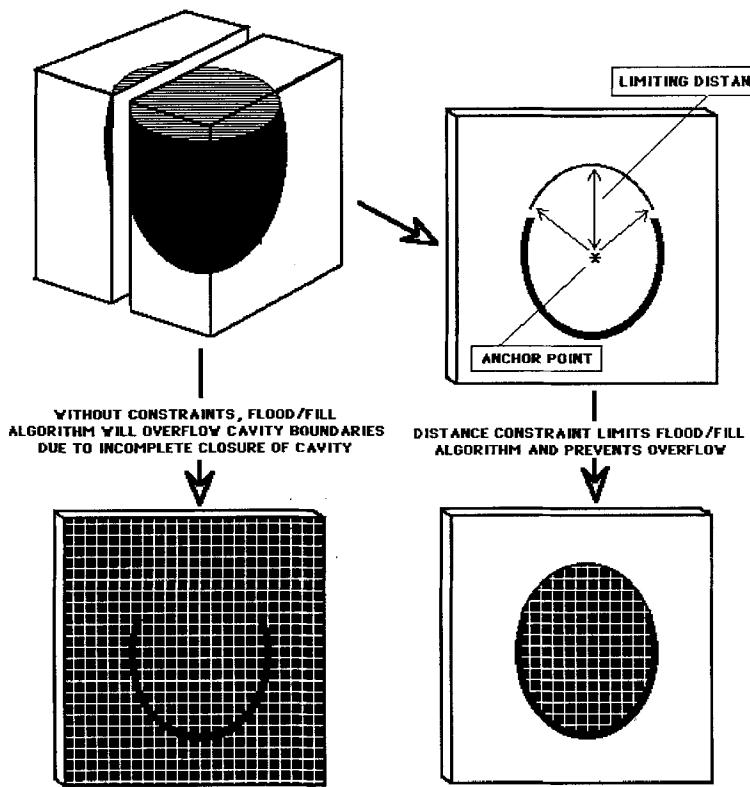


Fig. 7. The use of distance constraints to prevent overflow of the flood/fill search algorithm.

undate the cavity interior and spill out of the unbounded region. Therefore, a distance constraint is needed to contain the search process. This constraint is a limiting distance from some point within the cavity interior. If the atomic coordinates of a ligand or substrate molecule are known, distance constraints from these atoms are used. The flood/fill algorithm will then proceed from the seed point, but will be contained by the distance constraint.

We have shown that we can reconstruct the cavity interior one portion at a time by partitioning the cavity into slices. In practice, the cast is not reconstructed this way. Rather, as the filler lattice atoms are determined for each cross-section, their loci are simply stored in an array. Upon completion of the cavity search, a three-dimensional lattice of filler atoms results. It is this three-dimensional lattice of filler atoms that is transformed into the filler solid. Thus, only a single Boolean operation is necessary as the volume of the cavity wall atoms is subtracted from the filler solid, resulting in a perfect reproduction of the cavity interior.

As stated earlier, molecular surface displays may be color-coded to depict electrostatic potentials [18]. This can also be done with our cavity display. Having calculated the loci of filler atoms necessary to pack the cavity, we focus our attention upon those present within the outermost layer of the filler solid. They essentially form the lining of the cavity. In other words, these points lie along the *cavity-pocket interface* and are positioned where electrostatic interactions between the

pocket and the binding ligand may be measured. At each of these positions, the electrostatic potential of the atoms forming the cavity is calculated using the method described by Weiner et al. [5, 6] to compute the electrostatic potential at a specific point in space within a system of charges. These values are then scaled, assigned a color, and displayed.

Researchers interested in studying the electrostatic *interactions* between binding molecular entities usually do so by color-coding the molecular surface of each molecule by electrostatic potential. These surfaces are then docked and visually inspected to note regions of electrostatic complementarity and disparity. Although this method is quite effective, it requires the viewer to scrutinize both electrostatic color surfaces and mentally estimate the degree of electrostatic attraction and repulsion. To simplify displays of this type, we have color-coded the *electrostatic complementarity* between binding molecular entities. At every cavity-pocket interface point, the electrostatic potential of *both* the atoms forming the cavity and those of the binding ligand are calculated. A rough approximation of complementarity is computed by multiplying these potentials together. A favorable electrostatic interaction is produced when the electrostatic potentials are opposite in sign. Therefore, *favorable* interactions are indicated when the product of these values is a *negative* number. Likewise, *unfavorable* interactions are indicated when the product of these values is a *positive* number since the potential of the cavity and that of the binding ligand should have the same sign. These products are then scaled, assigned a color, and displayed.

Another useful application is the determination of regions where free space is present within the cavity when a ligand is bound. Such knowledge is very helpful in the design or modification of ligands. To achieve this end, a density mapping algorithm similar to that described above is used. A three-dimensional grid is employed to calculate a density map where each data point is a function of the distance to the nearest atom. In this way, the coordinates of regions where the distance to the nearest atom is greater than the radius of a specified target atom size may be computed. Calculations involving the computation of maps of this sort are often hindered by limitations in both memory and time, as both of these are directly related to the cavity size, number of atoms, and search grid density. To reduce this problem, we limit and concentrate our search efforts to regions centered about specific atomic loci. In order to investigate thoroughly where modifications to the bound ligand may be made, regions centered about the coordinates of the filler atoms located at the cavity-pocket interface should be searched. It is here that gaps between the ligand and the cavity exist. The locations where free space is found are displayed using filler spheres.

Although this free space visualization algorithm is useful for locating selected regions where ligand modifications may be made, it is not well suited for fully characterizing the void that exists between the ligand and the receptor, the ligand-receptor gap region. This technique merely shows the locations where an entity of a particular size may be placed; no information concerning the relative dimensions of free space is revealed. To facilitate the display of this information, we developed another algorithm to color-code the cavity display by *ligand-receptor nearest atom gap distance*.

Figure 8 depicts how this measure is computed. We measure this distance at each lattice point lying along the cavity-pocket interface. From a given lattice point, the closest ligand atom is first found. A vector is then generated from the ligand atom to the lattice point and extended into the space occupied by the receptor. All receptor atoms lying within a specified target distance to this vector are found, and the closest one to the ligand atom is selected. The actual *van der Waals surface to surface* distance (not center to center) between the ligand and receptor atoms is calculated.

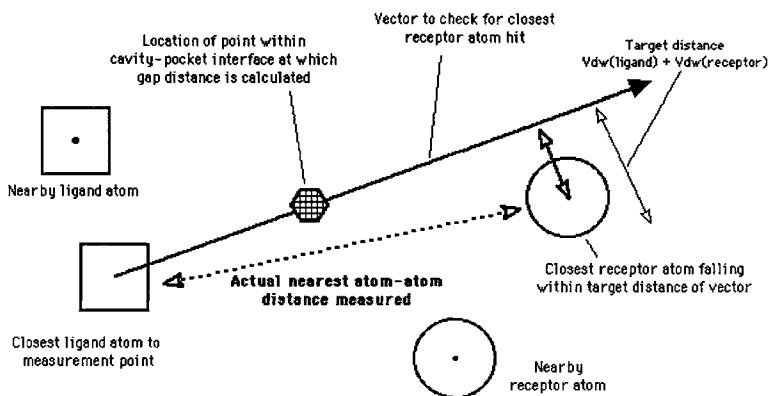


Fig. 8. Calculation of ligand-receptor nearest atom distance.

If no receptor atoms are hit by the vector, it is assumed that only solvent surrounds the ligand at that point. If this is the case, a default maximum distance value is assigned and the next lattice point is processed. Otherwise, to insure that a unique arrangement of atoms does not cause an erroneous measure, the above process is repeated except that the vector is instead calculated from the nearest receptor atom and extended into the space occupied by the ligand. The shorter of the two atom-atom distances is assigned to that lattice point, and the process continues. When the ligand-receptor distances have been calculated at all cavity-pocket interface lattice points, a user-defined color-coding scale is implemented to generate the displays.

COMPUTATIONAL METHODS

X-ray crystallographic background

The enzyme-substrate system used to generate the cavity displays was the structure of a complex between a chemically synthesized HIV-1 protease and a substrate-based inhibitor, MVT-101 [19]. The HIV-1 protease is a dimeric structure, with the active site located in a cavity-like region situated at the interface between the two subunits. Upon binding within this region, the inhibitor introduces substantial conformational changes to the enzyme. These changes may be described as hinge-like movements of the subunits which cause long, flap-like structures to close down upon the inhibitor, effectively capping the active site pocket. This action forms a true cavity, as nearly 80% of the inhibitor is excluded from contact with the surrounding solvent by the protein molecule. Individual side chains of the inhibitor itself are located in well-defined hydrophobic pockets. This structure was solved at 2.3 Å resolution, and the molecular features of the complex afforded us an excellent opportunity to demonstrate our new cavity-displaying algorithms.

Software development

All algorithms were implemented in FORTRAN on a VaxStation 3520. Solid modeling calculations were performed using the molecular modeling program SYBYL 5.3 installed on a Silicon Graphics IRIS-4d/80. All displays were created using SYBYL 5.3 and viewed with an Evans and

Sutherland PS390. At this time, the Cavity-Search software is compatible only with SYBYL and is available upon request from the Center for Molecular Design.

Creation of HIV-1 cavity display

PDB files containing the molecular coordinates of the HIV-1 protease (3124 atoms) and the MVT-101 inhibitor (117 atoms) were used as input. The {x,y,z} coordinates of the α -carbon of the third residue of the inhibitor, norleucine, were used as the seed location for the cavity search program. A search grid of 0.3 Å step size was used to fill the cavity space and a distance constraint of 2.5 Å from the inhibitor was used to limit the filler lattice to the active site region. Approximately 7 min of CPU time (Vax 3520) were necessary to create a file containing the coordinates of the filler atoms necessary to create the filler solid. The SYBYL molecular modeling program was then used to perform the solid modeling operations necessary to produce a graphical display file of the cavity cast.

Creation of electrostatic potential display

Electronic charges were assigned to each atom in the protease using parameters from the AMBER force field derived at UCSF [5, 6]. The electrostatic potential was calculated at the {x,y,z} coordinate location of each solid filler point lining the cavity-pocket interface (approximately 9500 individual loci). The cutoff distance for disregarding electrostatic contributions to the calculation was set at 8 Å. Computed potential values were rescaled and assigned a representative color. 'Hotter' colors (red, orange, yellow, and white) were used to denote regions of positive potential, with red being the most positive. 'Cooler' colors (cyan, blue, magenta, and violet) represented regions of negative potential, with violet being the most negative. Approximately 10 min of CPU time (Vax 3520) were required for the calculation. Displays were generated and viewed using SYBYL 5.3.

Creation of electrostatic complementarity display

Electronic charges were assigned to each atom in the protease and inhibitor using parameters from the AMBER force field. Using the algorithm described above, the electrostatic complementarity was calculated at the {x,y,z} coordinate location of each solid filler point lining the cavity-

TABLE I
GAP DISTANCE COLOR CODE

Color	vdW surface to surface distance (Å)	Comments
Red	0.00–0.95	No room present for any atom type
Orange	0.95–1.10	Minimum room for hydrogen
Yellow	1.10–1.40	Minimum room for oxygen
Cyan	1.40–1.65	Minimum room for carbon and nitrogen
Blue	1.65–3.00	Minimum room for sulfur, iodine, and other atoms of similar size
Purple	3.00–infinite	Regions surrounded only by solvent

vdW: van der Waals

pocket interface (approximately 9500 individual loci). The cutoff distance for disregarding electrostatic contributions to the calculation was set at 8 Å. Computed complementarity values were rescaled and assigned a representative color. ‘Hotter’ colors (red, orange, yellow, and white) were used to denote regions of electrostatic disparity, while ‘cooler’ colors (cyan, blue, magenta, and violet) represented regions of favorable electrostatic interactions. Approximately 10 min of CPU time (Vax 3520) were required. Displays were generated and viewed using SYBYL 5.3.

Creation of molecular free space display

Using the free space search algorithm described above, lattice points at the cavity-pocket interface were used as anchor positions to conduct searches for regions capable of accommodating a water molecule. Every anchor point served as the center of a $2 \times 2 \times 2$ Å voxel, each of which was systematically searched using a step rate of 0.2 Å for loci whose minimum distance to any other atom was no less than 1.5 Å (an approximate radius of a water molecule). To display the results, spheres were positioned at loci satisfying the above conditions. Approximately 20 min of CPU time (Vax 3520) were necessary to complete the calculations.

Creation of nearest atom gap distance display

Each lattice point along the cavity-pocket interface was color-coded according to the nearest atom gap distance algorithm described above. Table 1 depicts the color coding scheme used. Approximately 10 min of CPU time (Vax 3520) were required. Displays were generated and viewed using SYBYL 5.3.

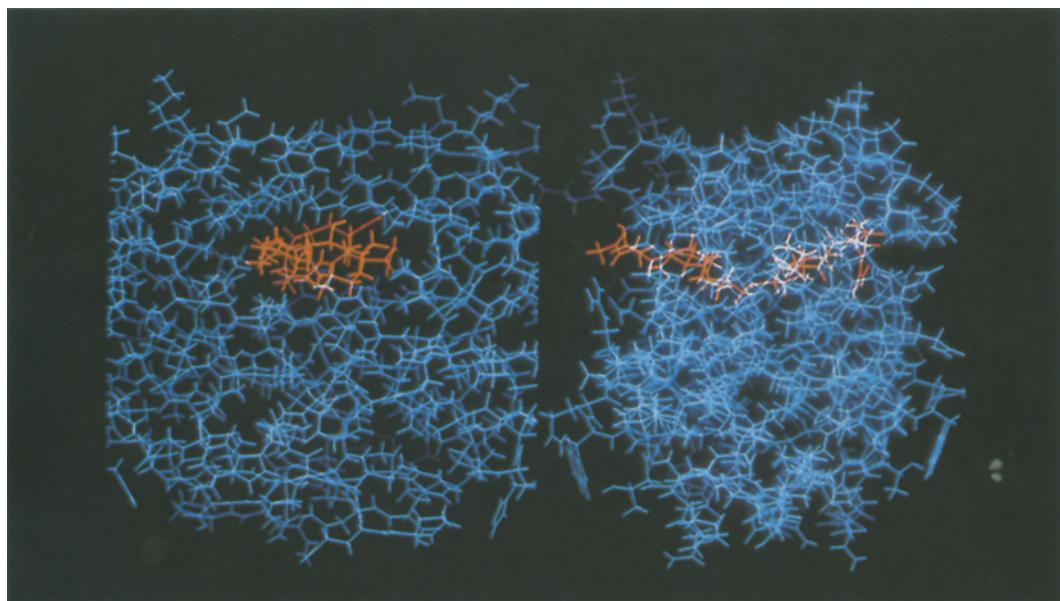


Fig. 9. Orthogonal views of MVT-101 inhibitor (red) shown bound within the active site of HIV-1 protease (blue).

RESULTS

The MVT-101 inhibitor is shown in Fig. 9 bound within the active site of the HIV-1 protease. The inhibitor is nestled within a cavity-like region formed by the interface of the two protease subunits. Two flap-shaped structures lie over the opening of this cavity, enclosing the inhibitor within. This configuration surrounds the inhibitor in its entirety, leaving only the ends exposed to the solvent. The atoms making up this binding cavity belong to several different domains, and it is readily apparent in this stick model that the structures involved are quite complex in shape. It is this structural complexity that we wish to reduce by implementing our cavity-displaying techniques.

Figure 10 shows the cast of the cavity volume as it sits in relation to the enzyme. Since the cast was created using solid modeling Boolean functions, it is an absolute replica of the internal volume of the cavity. There are numerous indentations imparted upon the cast by the atoms of the protease making up the cavity wall; however, in the regions exposed to the solvent, the cast surface is smooth because the filler solid was contained solely by distance constraints.

The use of clipping mechanisms to graphically slice through the cast offers the most useful views of the cavity. As shown in Fig. 11, when the cast is sectioned, the view one sees is not that of a cast, but rather a *shell*. This shell runs along the lining of the pocket, adhering strictly to the molecular surface of the atoms defining the cavity wall. Note that this is *not* a Connolly solvent-accessible surface, but is instead the actual van der Waals surface. In Fig. 12, dot surfaces of the inhibitor atoms have been included. Steric relationships can easily be seen and understood. Note that the

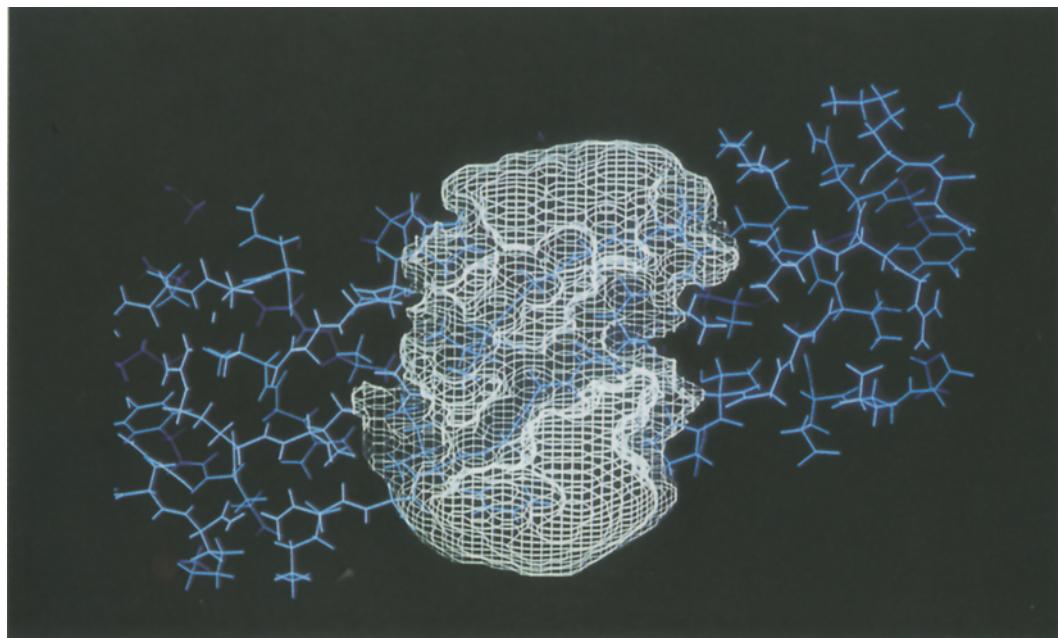


Fig. 10. External view of solid model cast shown bound within active site region of HIV-1 protease.

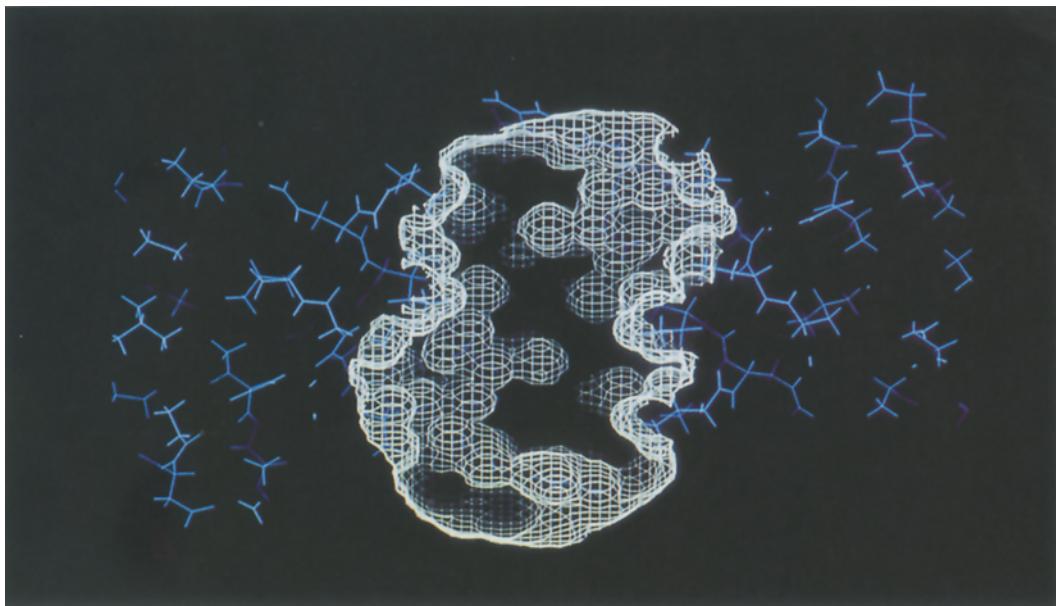


Fig. 11. Sectioned view of HIV-1 protease active site cast showing cavity shell lining the van der Waals surface of cavity wall atoms.

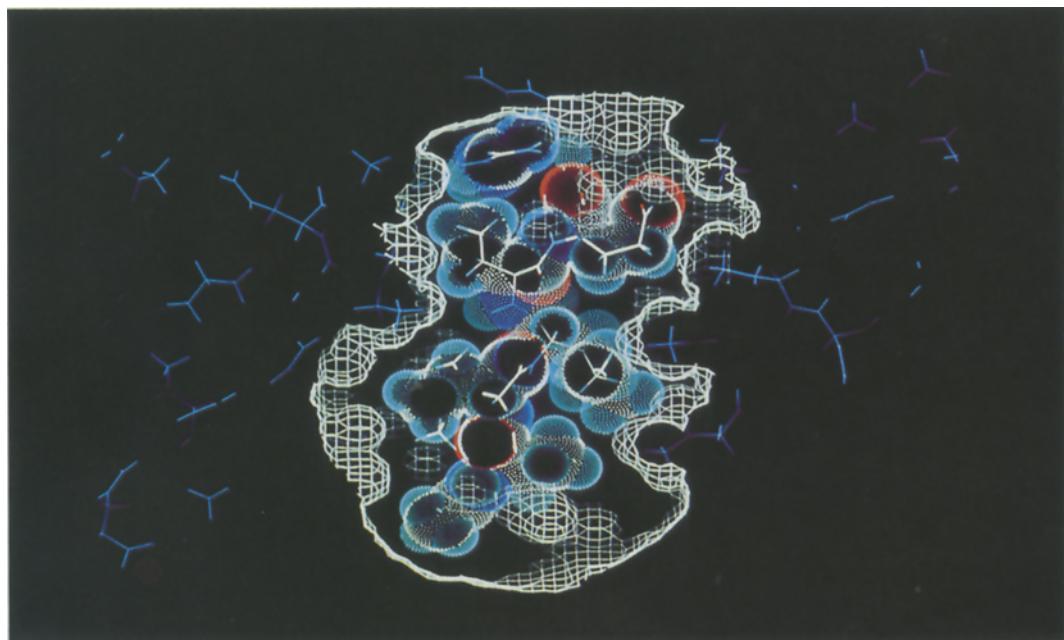


Fig. 12. Steric complementarity between van der Waals surface of ligand atoms and cavity shell.

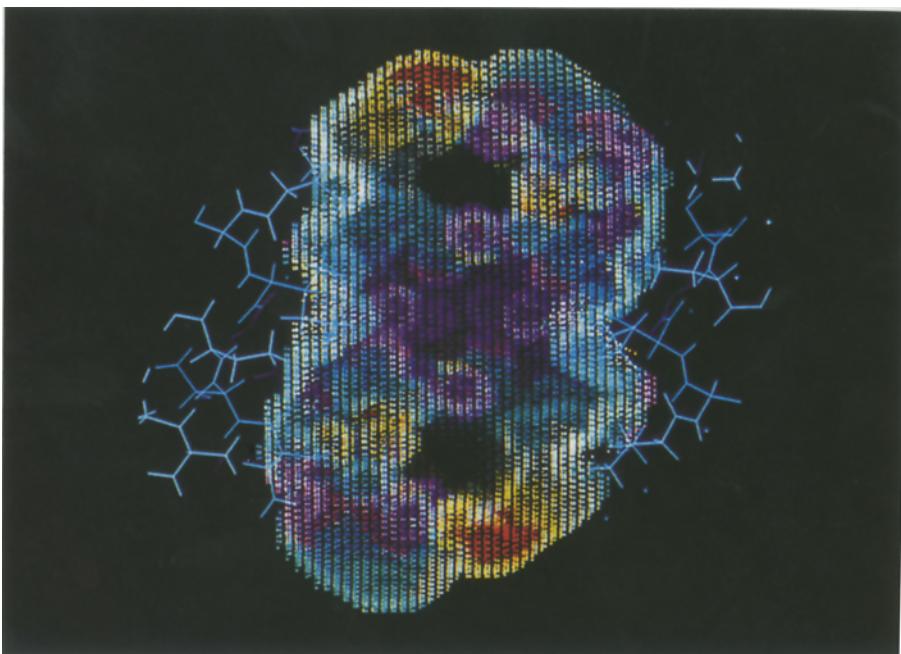


Fig. 13. Electrostatic potential of HIV-1 protease shown painted upon inner surface of cavity shell. Red (highest positive potential) > orange > yellow > white > cyan > blue > magenta > violet (lowest negative potential).

atoms of the protease need not be included, as the ‘molecular layout’ of the cavity is faithfully represented. Thus, the full attention of the investigator may be applied to the study of the binding region.

In addition to steric constraints, electrostatic interactions play a crucial role in the binding affinity between molecular entities. Figure 13 depicts the *electrostatic potential* of the protease painted upon the inner surface of the shell. With this display, the entire electrostatic environment surrounding the inhibitor may be characterized.

As explained above, with both the enzyme and the inhibitor present, we can simply calculate the *electrostatic complementarity* between the two interacting species. This function is then color-coded and ‘painted’ upon the cavity shell. Figure 14 displays several hydrogen bonds that form when the inhibitor is bound within the active site. The identical scene is reproduced in Fig. 15, except that the shell and its painted surface replace the atoms of the cavity. The color coding is spectral, with ‘hotter’ colors (red, orange, yellow, and white) representing electrostatic disparity and ‘cooler’ colors (cyan, blue, magenta, and violet) indicating favorable interactions. As one can see, the hydrogen bonds pierce the shell precisely and specifically within violet-colored areas, indicating that these are indeed regions of most favorable electrostatic interaction. By displaying *complementarity* rather than electrostatic potentials, one can immediately discern which regions of the inhibitor are beneficial to binding and which segments may have to be modified to improve the inhibitor.

In order to modify portions of a binding ligand, one must have a good understanding of where

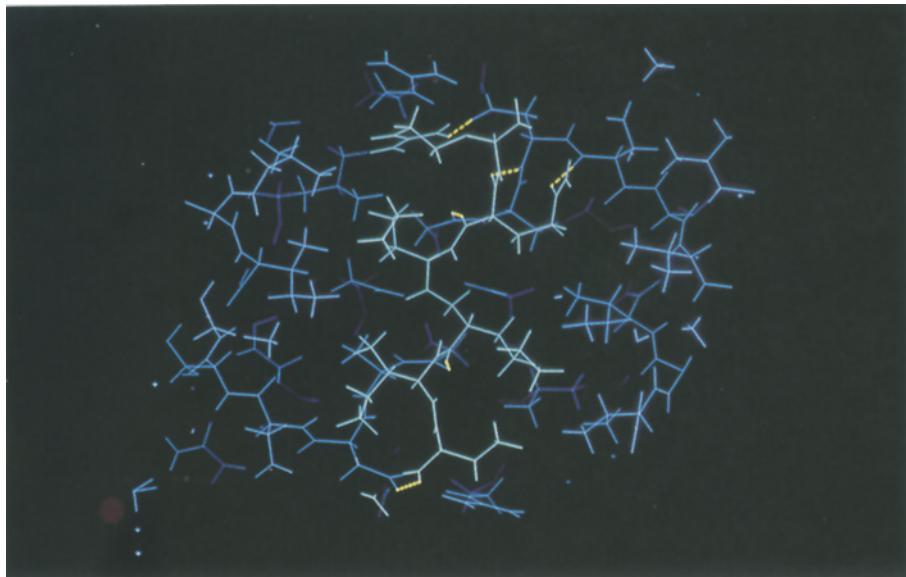


Fig 14. Hydrogen bonding interactions (yellow) between MVT-101 and HIV-1 protease.

free space exists when the ligand is bound within the cavity. Figure 16 depicts regions of space surrounding the inhibitor that are large enough to accommodate a water molecule when the inhibitor is bound within the cavity. Van der Waals surfaces of water molecules actually present within the crystal structure have been included in red to verify the position of these volumes. These regions were found using the free space searching algorithm described above. The darker colored volumes

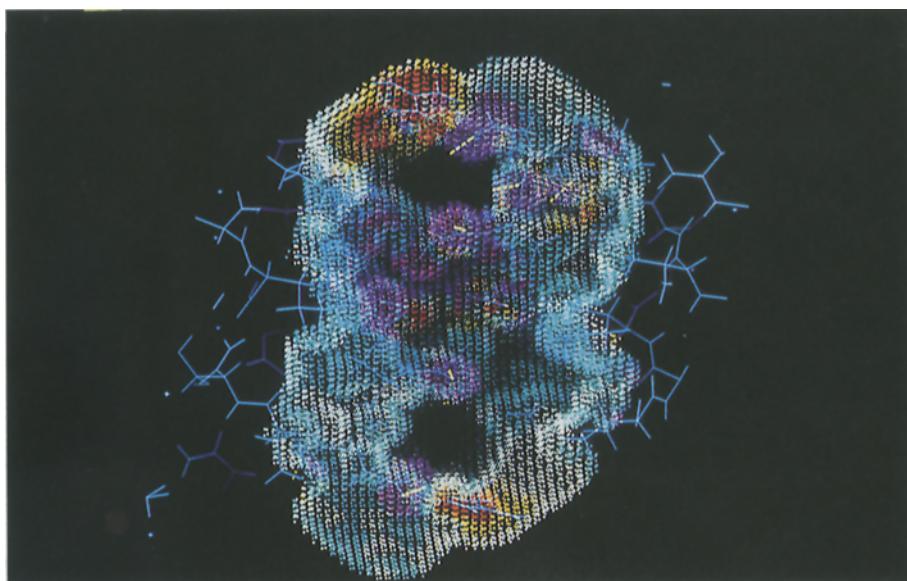


Fig. 15. Electrostatic complementarity between HIV-1 protease and MVT-101 shown painted upon inner surface of cavity shell. Red (greatest disparity) > orange > yellow > white > cyan > blue > magenta > violet (greatest complementarity).

differ from the lighter colored ones in that they are regions of space that primarily extend into solvent as opposed to being surrounded by protein. Thus, in Fig. 16, the dark blue volumes are located at the open ends of the cavity, whereas the cyan-colored volumes represent areas that are buried within the protease.

To fully characterize the space between the protease and the inhibitor, we can also color the shell by nearest-atom gap distance. Figure 17 depicts the results of such a calculation, and Table 1 lists the color-coding scheme used for this display. The majority of the shell is colored red, indicating that there is indeed a tight fit between the inhibitor and the enzyme. By setting appropriate distances for the color-coding scheme, one can immediately discern where chemical groups of various sizes may be used to replace or modify other sites on the inhibitor. As an example, suppose that the region surrounding a particular side chain of the inhibitor is colored yellow. With our color scheme, this would indicate that there is physically enough room to pack a nonbonded oxygen or hydrogen between the van der Waals surfaces of the inhibitor and the enzyme. Since bonded atomic distances are shorter than nonbonded distances, we can confidently add an oxygen or hydrogen to the side chain without fear of steric contact. Note that the regions of the inhibitor in Fig. 17 that are exposed solely to solvent are colored dark blue to violet.

CONCLUSIONS

With the development of these novel algorithms and techniques, we have created tools with which an investigator can more effectively study and understand the interactions taking place between a cavity-like binding site and a complementary ligand. Computer graphics and imaging

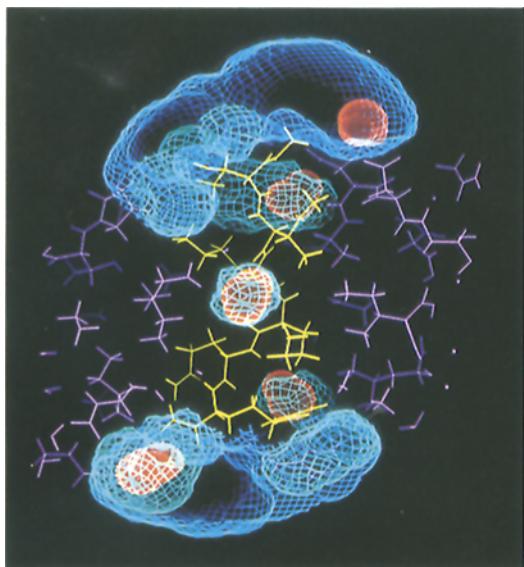


Fig. 16. Regions within the protease-inhibitor complex large enough to accommodate a bound water molecule. Water molecules actually present within crystal structure shown depicted in red. Dark blue volumes represent regions extending into solvent, while cyan volumes represent regions surrounded by the enzyme.

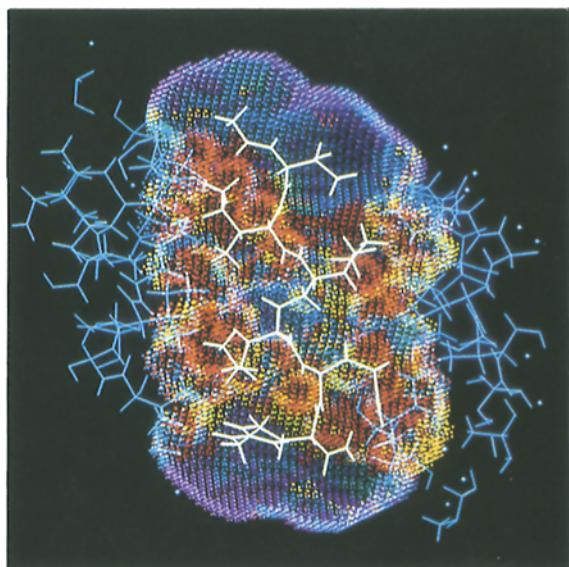


Fig. 17. Inner surface of cavity shell shown painted by ligand-receptor nearest atom gap distance between MVT-101 (white) and HIV-1 protease active site atoms (blue). See Table I for color-coding scheme.

technology have advanced to the point where an overwhelming amount of information may be displayed at the investigator's disposal. The current challenge is to limit and focus such information as efficiently as possible. As is often the case, less is more if the less is relevant.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (grant GM24483) and the Medical Scientist Training Program in the Division of Biology and Biomedical Sciences at Washington University (training grant GM07200).

REFERENCES

- 1 Marshall, G.R. and Naylor, C.B., In Ramsden, C.A. (Ed.), *Comprehensive Medicinal Chemistry*, Vol. 4, Pergamon Press, Oxford, 1990, pp. 431–458.
- 2 Blaney, J.M. and Hansch, C., In Ramsden, C.A. (Ed.), *Comprehensive Medicinal Chemistry*, Vol. 4, Pergamon Press, Oxford, 1990, pp. 459–496.
- 3 Connolly, M.L., *Science*, 221 (1983) 709.
- 4 Richards, F.M., *Annu. Rev. Biophys. Bioeng.*, 6 (1977) 151.
- 5 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P.K., *J. Am. Chem. Soc.*, 106 (1984) 765.
- 6 Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., *J. Comput. Chem.*, 7 (1986) 230.
- 7 Fauchere, J.L., Quarendon, P. and Kaetterer, L., *J. Mol. Graphics*, 6 (1988) 203.
- 8 Furet, P., Sele, A. and Cohen, N.C., *J. Mol. Graphics*, 6 (1988) 182.
- 9 Fersht, A., *Enzyme Structure and Mechanism*, Freeman, New York, 1977.
- 10 Kuntz, I.D., Blaney, J.M., Oately, S.J., Langridge, R. and Ferrin, T.E., *J. Mol. Biol.*, 161 (1982) 269.
- 11 Voorintholt, R., Kosters, M.T., Vegter, G., Vreind, G. and Hol, W.G.J., *J. Mol. Graphics*, 7 (1989) 243.
- 12 Casale, M.S. and Stanton, E.L., *IEEE Computer Graphics and Applications*, 5 (1985) 45.
- 13 Requicha, A.A.G. and Voelcker, H.B., *IEEE Computer Graphics and Applications*, 2 (1982) 9.
- 14 Requicha, A.A.G., *IEEE Computer Graphics and Applications*, 3 (1983) 25.
- 15 Hibert, M.F., Hoffmann, R., Miller, R.C. and Carr, A.A., *J. Med. Chem.*, 33 (1990) 1594.
- 16 Sufrin, J.R., Dunn, D.A. and Marshall, G.R., *Mol. Pharmacol.*, 19 (1981) 307.
- 17 Foley, J.D. and Van Dam, A., *Fundamentals of Interactive Computer Graphics*, Addison-Wesley, Reading, MA, 1982.
- 18 Weiner, P.K., Landridge, C., Blaney, J.M., Schaefer, R. and Kollman, P.A., *Proc. Natl. Acad. Sci. U.S.A.*, 79 (1982) 3754.
- 19 Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, A., *Science*, 246 (1989) 1149.