The use of an algorithmic method for small molecule superimpositions in the design of antiviral agents

Guy Diana, Edward P. Jaeger, Melissa L. Peterson and Adi M. Treasurywala*

Sterling Winthrop Pharmaceuticals Research Division, 81 Columbia Turnpike, Rensselaer, NY 12144, U.S.A.

SUMMARY

The inability to reliably predict relative orientations of drug molecules within our series of antipicornavirus agents has severely limited the usefulness of available structure–activity data in the drug design process. A reported method of overlapping molecules has been evaluated to see if it could provide a solution to this problem. Although it initially succeeded remarkably well with a series of molecules whose bound X-ray structures were known, this success was shown to be only a function of the bound conformation of these molecules. Thus, this method did not provide a general solution to the problem at hand.

INTRODUCTION

The 3D structure of Human Rhinovirus 14 (HRV-14) was reported by Rossmann et al. in 1985 [1], adding much insight to the structure and function of this virus. Subsequently, X-ray crystallographic data of several antirhinovirus compounds bound to HRV-14 revealed that the compound binding site was a hydrophobic pocket located in VP1 of the capsid protein [2] in the vicinity of a 'canyon' which has been shown to be the cell receptor (Fig. 1) [3]. In addition, early studies on a series of related compounds demonstrated that these structures could bind in either of two orientations (Fig. 2); observations which were not understandable at the time [4]. As X-ray analyses of different members of this family of organic molecules bound in the same virus serotype were accumulated, it became apparent that different molecules bound (apparently quite specifically for each molecule) in one of two orientations which were related to one another by a 180° rotation of the long axis of the molecule (Fig. 2). No obvious relation could be found between the observed orientation as found by the X-ray analysis and any computable properties [4]. For example, a molecule that was observed to bind in one orientation could be flexibly docked

---

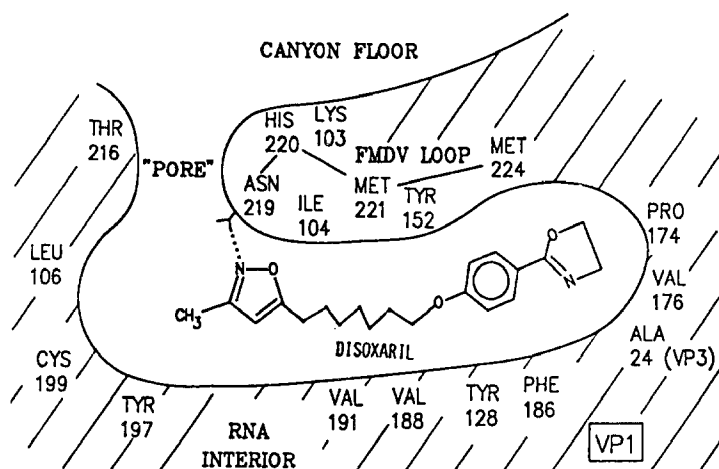*To whom correspondence should be addressed.

Fig. 1. Schematic of binding site for HRV-14.

into the same pocket in the opposite orientation with an overall difference in energy between the two possible states that was well within the error margins of the calculations (see below). This inability to correlate the binding mode with some clearly computable term meant that meaningful 3D structure–activity relations could not be developed for all the molecules that had been synthesized and tested against a given serotype. Only after the crystallographic analysis of the molecule bound in the virus capsid protein had been completed could the orientation be known with confidence and only then could such molecules be meaningfully overlaid to produce 3D SARs. This was indeed a considerable handicap since the number of molecules for which biological data were available was far in excess of the number for which X-ray structures could be obtained.

The problems involved with reliably predicting the orientation of a small molecule within a biopolymer based on the known crystal structure of another small molecule of a comparable size bound to the same macromolecule are not unique to HRV-14. Other well-known examples include dihydrofolate reductase [5] and streptavidin [6]. In the latter case the researchers performed computations based on a certain orientation of small molecules only to find out through crystallographic studies in each series that closely related small molecules had entirely different orientations within the biopolymers being studied. Indeed, the existence of multiple binding modes for the same or similar molecules is becoming a better documented event in general.

The analysis of previously reported crystal structures of several of our designed antiviral agents in the so-called WIN pocket of Human Rhinovirus 14 has yielded many interesting clues to the mechanism of action of these agents in preventing the uncoating of the virus particles with subsequent inhibition of viral replication [7]. To further our understanding of the SAR of this class of compounds as rhinovirus replication inhibitors, a method was sought that could reliably predict the orientation of a given molecule in the pocket of HRV-14. Such information would provide a better understanding of the entire body of SAR data at our disposal. If one were able to orient these molecules correctly without recourse to X-ray analysis of each and every one of them, then one could superimpose them and develop better 3D-QSAR approaches towards designing new molecules that were even better. Knowledge of the orientation of a compound in
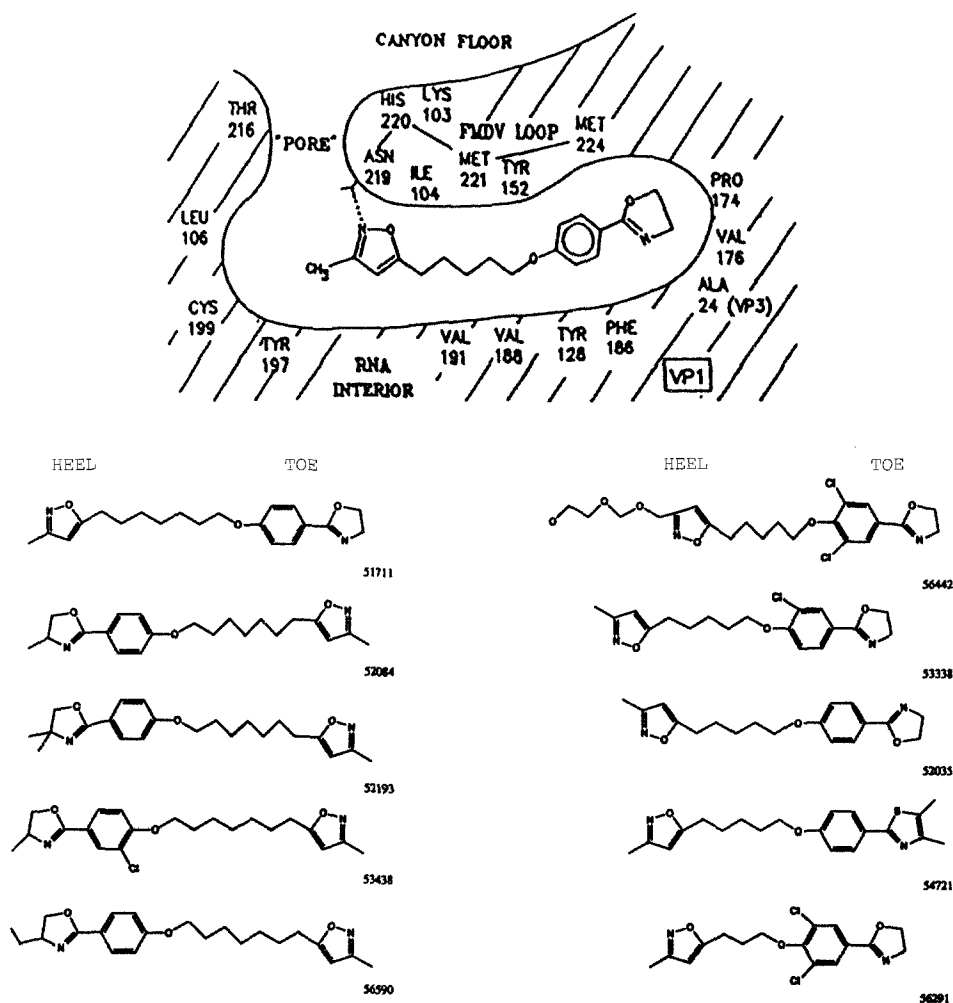
Fig. 2. Orientations of molecules used.

the pocket would permit us to distinguish between structural changes which increased binding affinity and those which alter binding characteristics such as orientation or binding mode.

The method of simulated annealing of two structures as rigid geometric bodies to determine the best overlap was recently reported by P.M. Dean et al. [8]. We reasoned that if one could translate the issue of predicting the orientation of a given analog inside the pocket to one of predicting the relative best overlap of two molecules to one another in isolation then our problem would, in effect, be simplified to comparing analogs of unknown orientation to those of known orientation. If this method worked it would allow us to orient these molecules relative to one another, even in cases where X-ray structures were unavailable. Thus the approach offered the hope that all of the known structural and biological data could be used to generate a better 3D SAR. However, it was recognized that the success of this method depended on two factors: (1) the chemical constitution (connectivity and functionalities) of the molecules being considered, and (2) the actual conformation of each of the two molecules which was chosen. The importance of the latter factor to the

outcome of this method had not received much attention in the past [8] and so we set out to examine this point further. It is the result of this approach to the problem that we wish to report here.

EXPERIMENTAL

All calculations were performed on a Sun SPARCserver 490 running SYBYL 5.41. Coordinates of the structures used in this study are available in SYBYL molfile format as supplemental material. The essential features of the simulated annealing algorithm have been incorporated into the command SUPERIMPOSE in SYBYL and it is this implementation that was used.

RESULTS

The experimental scheme used here was to apply the simulated annealing overlap method of Dean et al. to determine the best overlaps between pairs of molecules. Structure matching problems which focus on overlapping atom positions of 'corresponding' atoms between chosen conformers of two molecules may be considered at two levels.

(1) Fitting of single conformers of each molecule (rigid body fitting) where the atoms which should correspond to each other are specified. This problem reduces to the minimization of the
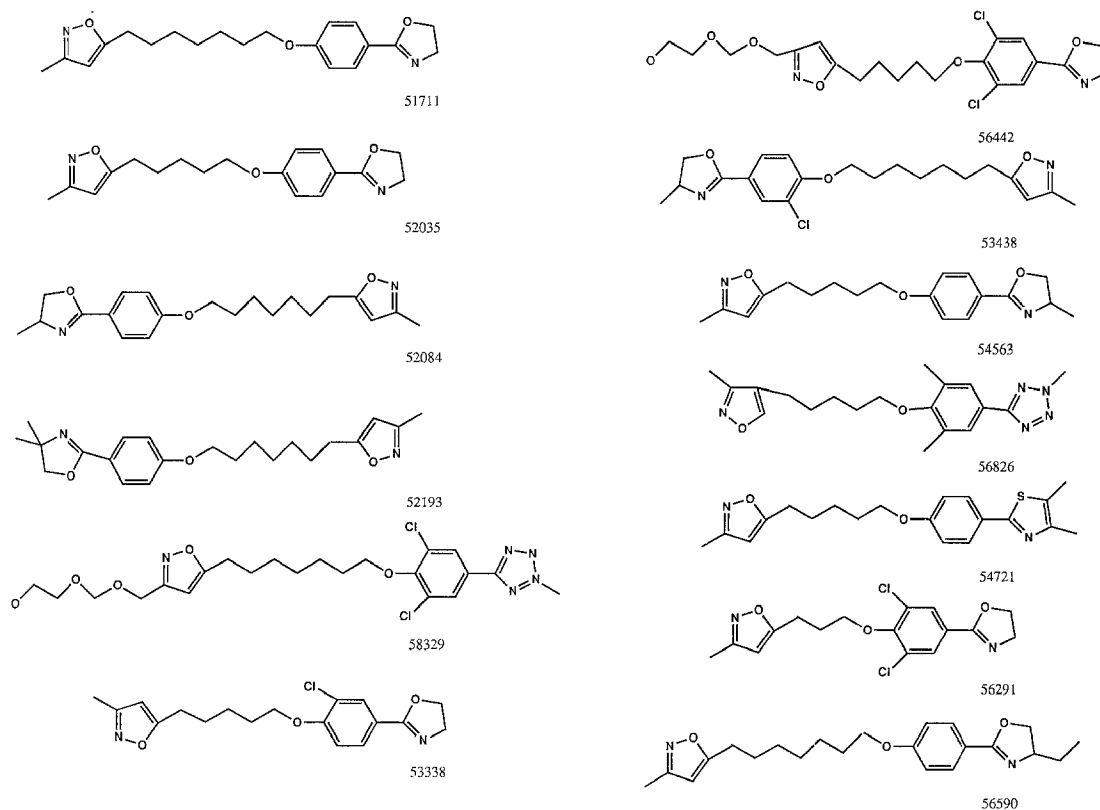

Fig. 3. Conformations of structures used.

difference between the corresponding coordinates by applying only a translation and/or a rotation step. This is a well-known and often used technique.

(2) Rigid body fitting of two molecules where there are no predetermined atom correspondences between dissimilar molecules, A and B, containing NA and NB atoms, respectively. The method of Dean et al. addresses how one would find this optimal set of atom correspondences such that the distance difference matrix of these pairs of atoms is minimized. Their method simply treats each atom as a point in space. No chemistry is used in overlapping any particular points (e.g. heteroatoms on other heteroatoms etc.). Once this set of correspondences is arrived at, one molecule is reorientated through appropriate translations and rotations to present the result visually.

The molecules that were chosen for the test phase of this study (Fig. 3) had all been analyzed by X-ray crystallography while bound to HRV-14. Thus their orientation as well as their so-called bioactive conformation was known. These molecules were extracted without minimization from the viral capsid structures reported by Rossmann et al. [9] (see below for an analysis of the effects of minimization on the conclusions drawn). These positionally and orientationally aligned structures formed the starting point of the study. It was felt that this initial set formed a good test of the method since the success of the algorithm could be directly evaluated against the real data without any doubt about the relevance of the conformation of each molecule used. Their computed relative orientation could then be extracted from the best overlaps, because the molecules are roughly linear in shape, and the best fits should be found in one orientation along this major axis or the other. This computed preference for a certain relative orientation can then be compared to the crystallographically observed orientation.

METHODOLOGY

The simulated annealing overlap method of Dean et al. works by selecting equivalently sized subsets of atoms from the two molecules to be overlapped. These subsets are then permuted to identify which atoms in the two subsets should be superimposed. Once the atoms are identified, the RMS fit of these permuted subsets is computed. The selection and permutation process is continued over a number of cycles and the best ten overlaps are reported.

All atoms of each molecule except hydrogens or halogens were considered for overlapping. These were included in the so-called list of 'interesting atoms'. The molecule containing the larger number of atoms of interest was designated as the 'template'. The second molecule was reduced to a list of interesting atoms (i.e. all atoms except hydrogens and halogens) and an attempt was made to fit all of the 'interesting' atoms of this second, so-called 'target' molecule to an equal number of randomly selected atoms of the template molecule. The atom pairing between the template and target atoms was repeatedly permuted for a given selection (the number of repeats was set here at 100). The first match that resulted in an overall RMS value for all of these pairwise overlaps that was less than that obtained for the previous sets of atom choices was retained. If no such match was found within 100 permutations then a new set of atoms was chosen from the template. This is referred to as the next selection. Altogether, the process continued until 1000 valid trials had been accumulated. The output of this algorithm was the set of the ten best RMS overlaps between the template molecule and the target molecule determined in this way. A flow chart of this algorithm is shown in pseudo-code below.

Flow chart:
    generate template atom list
      (list of all atoms *minus* all hydrogens, or halogens)
    generate target atom list
      (list of all atoms *minus* all hydrogens, or halogens)
      for 1000 selections
        {select template atoms for subset list (equal in member to the entire target atom list)
         for 100 permutations
          {permute the template atom subset list
          compute the RMS between the atoms in
            target atom list and the atoms
            in the template atom subset list
          save RMS and superimposition in the set of
            10 best RMS' and superimpositions
            if better than previous set of atom choices
         }
        }
    report 10 best RMS' and corresponding superimpositions

From this set of the top ten superimpositions of two molecules, the relative 'computed' orientations of the two molecules for each of these ten fits were recorded. These results were then compared with the relative orientation of the same two molecules as obtained from the crystal structures: the so-called 'observed' orientation. These data are recorded in Table 1. If the two molecules were both observed with the isoxazole ring deep in the 'toe' (see Fig. 1) of the pocket in their respective X-ray structures then the relative 'observed' orientation was designated 'head-to-head' and marked by a '+' sign in the table. If they were observed to be in *opposite* orientations to one another then they were designated as 'head-to-tail' and marked with a '−' sign in Table 1. Fig. 4 contains an example of a 'head-to-head' and a 'head-to-tail' orientation. The 'computed' orientation for a set of superimpositions is measured by counting the number of times that the two molecules were oriented in the 'observed' way in the ten top fits. As an example, WIN53438 and WIN56590 were observed 'head-to-head' in their X-ray structures and this is exactly what the algorithm predicted: ten out of ten times. The observed orientation is marked with a '+' in Table 1 and thus the computed orientations are summarized with a '+10'.

All 13 molecules were compared to each other in this pairwise fashion. One of the ten best superimpositions found by the algorithm for WIN52035 and WIN52084 is shown in Fig. 5 which indicates how the method can find overlaps which are not intuitively obvious (the phenyl rings are not superimposed) for example. For this pair of molecules, the algorithm found the head-to-tail orientation nine out of ten times and this orientation agrees well with the virus crystallographic results (−9).

As mentioned above the method is based on a random number approach to the selection of atoms in the target to be superimposed. To test the possibility that the correlations we were observing were chance ones, the entire study was repeated a second time. Although the exact numbers differed by one or two out of ten, the high degree of correlations that was observed in the first run was preserved in the second run.

TABLE 1
SUPERIMPOSED RESULTS OF X-RAY CRYSTALLOGRAPHIC STRUCTURES

	56590	51711	53438	56826	52035	52084	56442	52193	53338	58329	54721	54563	56291
56590	+10	−9	+10	−6	−1	+1	−0	+1	−0	−1	−3	−0	−3
51711	−9	+10	−10	+8	+0	−0	+3	−0	+1	+3	+10	+1	+9
53438	+10	−10	+10	−4	−2	+0	−0	+2	−1	−1	−1	−2	−3
56826	−6	+8	−4	+10	+3	−0	+3	−0	+5	+4	+10	+7	+7
52035	−1	+0	−2	+3	+10	−9	+10	−8	+10	+7	+4	+4	+6
52084	+1	−0	+0	−0	−9	+10	−10	+10	−8	−7	−7	−5	−1
56442	−0	+3	−0	+3	+10	−10	+10	−10	+7	+10	+6	+9	+5
52193	+1	−0	+2	−0	−8	+10	−10	+10	−8	−10	−6	−6	−4
53338	−0	+1	−1	+5	+10	−8	+7	−8	+10	+6	+4	+7	+4
58329	−1	+3	−1	+4	+7	−7	+10	−10	+6	+10	+8	+8	+7
54721	−3	+10	−1	+10	+4	−7	+6	−6	+4	+8	+10	+6	+3
54563	−0	+1	−2	+7	+4	−5	+9	−6	+7	+8	+6	+10	+4
56291	−3	+9	−3	+7	+6	−1	+5	−4	+4	+7	+3	+4	+10

In a subsequent analysis (suggested by one of the referees) the entire matrix of comparisons (see Table 2 below) was repeated after each molecule had been minimized to complete convergence in the Tripos force field (TFF 5.41) [10] (see for example Fig. 6). Some changes were observed in the actual numbers obtained but they appeared to be randomly distributed throughout the matrix and significant changes (more than a score change of 3) occurred in only nine out of 169 compari-

Fig. 4. A. Example of head-to-head structures (dashed line molecule = 52084, solid line molecule = 52035). B. Example of head-to-tail structures (dashed line molecule = 52193, solid line molecule = 53338).
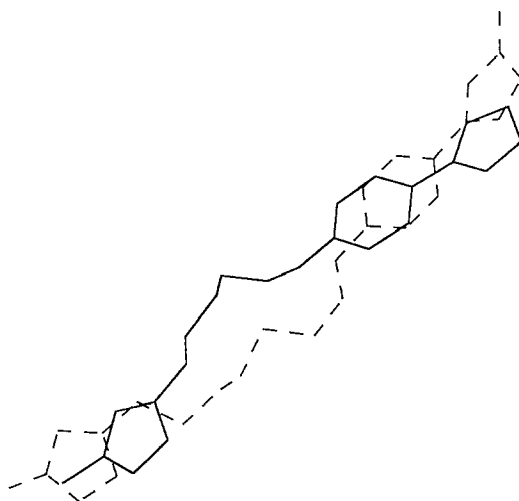
Fig. 5. One of the ten best superimpositions found by the algorithm for WIN52035 and WIN52084.

sons or in about 5% of the table. These results demonstrate that minimization does not significantly change the conclusion one can draw from this study.

It was clear that the algorithm was indeed selecting some real preference for different molecules to superimpose together that was somehow correlated to their relative orientations in the pocket. However, what was not tested so far in this study, or indeed reported in the literature to date [8], was the possible dependence of this observation on the starting conformation of the molecules. It was, in fact, reasonable to assume that the pocket-bound conformations that we were studying were carrying with them some 'imprint' of the pocket and that it was the superimposition of this imprint that was biasing the results in favor of the correct answer. To address this issue, we repeated this study once again. This time 2D connectivity data were used for each molecule as the

TABLE 2
SUPERIMPOSED RESULTS OF MINIMIZED X-RAY CRYSTALLOGRAPHIC STRUCTURES

|	56590	51711	53438	56826	52035	52084	56442	52193	53338	58329	54721	54563	56291
56590	+10	−10	+10	−4	−0	+1	−0	+0	−1	−1	−1	−0	−3
51711	−10	+10	−10	+10	+0	−0	+0	−1	+0	+2	+9	+2	+6
53438	+10	−10	+10	−8	−0	+0	−0	+3	−0	−2	−1	−0	−4
56826	−4	+10	−8	+10	+5	−0	+4	−1	+3	+3	+6	+5	+6
52035	−0	+0	−0	+5	+10	−9	+8	−8	+10	+8	+7	+8	+3
52084	+1	−0	+0	−0	−9	+10	−10	+10	−10	−9	−3	−4	−2
56442	−0	+0	−0	+4	+8	−10	+10	−9	+8	+9	+6	+8	+3
52193	+0	−1	+3	−1	−8	+10	−9	+10	−9	−9	−4	−4	−2
53338	−1	+0	−0	+3	+10	−10	+8	−9	+10	+10	+4	+8	+5
58329	−1	+2	−2	+3	+8	−9	+9	−9	+10	+10	+7	+9	+5
54721	−1	+9	−1	+6	+7	−3	+6	−4	+4	+7	+10	+6	+4
54563	−0	+2	−0	+5	+8	−4	+8	−4	+8	+9	+6	+10	+5
56291	−3	+6	−4	+6	+3	−2	+3	−2	+5	+5	+4	+5	+10
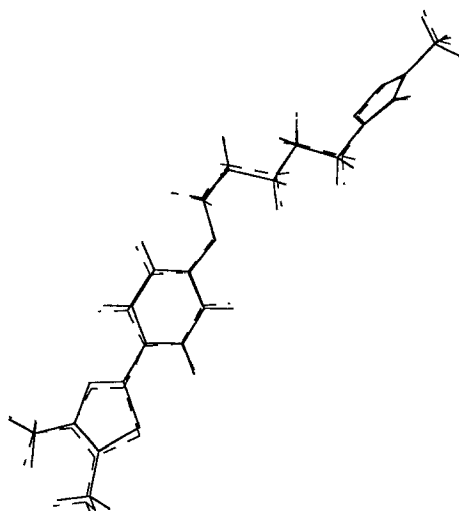
Fig. 6. Diagram showing the extent of movement in WIN54721 as a consequence of minimizing it in SYBYL 5.5 (Total RMS of all non-H atoms was 0.2455). Energies before and after minimization were 61.464 kcal/mol and 25.416 kcal/mol, respectively.

starting point (MDL's MACCS DBMS was the product used). These structures were all imported into the commercial molecular modeling package CHEMX [11]. This package has the facility that it implicitly translates the 2D structures into 3D structures via a rule-based algorithm. The 3D representations of these molecules were all found to be 'reasonable' since energy minimization (in TFF 5.41) did not result in a substantial movement of the atoms. The resulting structures were subjected to the same superimposition test using the simulated annealing algorithm. As seen from Table 3, this time the method failed to predict any orientational differences at all. They were all predicted to overlap head-to-head only.

DISCUSSION

From the results in Table 3 it can be inferred that the reason for the initial success was indeed that the molecules were manifesting the imprint of the pocket in their 'pocket-bound' conformations. This is an important finding since it demonstrates that there is a chiral requirement of overall conformation for binding. It also suggests that the information about the pocket can be extracted from the knowledge of the binding conformation of these molecules. This may point to a critical set of torsional angle values that must correlate with the activity of a given molecule. This finding in itself is an important one because it may guide the synthesis of compounds where these torsional values are even more strongly preferred. This aspect is currently under investigation.

However, the finding has some more important and general consequences. It has underlined the complete dependence of the final outcome of this method of overlapping and hence predicting orientation on the conformations of the molecules which are chosen. It points to possible further improvements in this method where the flexible matching of molecules in the same way can be addressed. The dependence of this algorithm on the starting conformation of the molecules has

not been considered explicitly before. However, in retrospect it is not surprising. The initial goal of finding a method that would reliably predict the orientation of this set of molecules in the pocket of HRV-14 remains an elusive one. It is perhaps noteworthly to point out again that energy calculations for the set of molecules in the pocket in the two possible orientations (relative to the long axis of the drug molecules) also proved not to be a reliable enough method of distinguishing between these choices for a given molecule as mentioned above. For example, the molecule WIN54721 was optimized inside HRV-14 in the reported orientation. The molecule was then frozen in its final geometry and removed from the virus. Its energy in isolation was found to be 25.455 kcal/mol. Then the whole molecule was turned head-to-tail and conformationally analyzed in the pocket to find the best conformation for the new orientation. This conformation was optimized in the pocket, extracted and then evaluated in a completely analogous fashion to the first. It was found to have an energy of 24.921 kcal/mol. Furthermore, these two conformations had very similar energies inside the pocket as well. This demonstrates that the conformational energy alone is not a sufficient criterion for predicting orientation in the HRV-14 pocket.

Several other conclusions can be drawn from these results. The first is that the data divides itself roughly into two groups. The correct orientation of members of one group is reliably predicted by other members of the same group. However, these compounds predict with almost *equal* reliability the *incorrect* answer for members of the second group. These groups have been delineated in Table 1. This is a fascinating result in itself. The exact reason for this type of behavior is not yet clear. It is worth bearing in mind that the resolution of these crystal structures is about 3–5 Å and minor differences in the exact conformation of the molecules may have a great influence on the final outcome of this analysis. The important point to consider is that these 'incorrect' predictions are no less certain than the 'correct' ones so that it is not a case of the algorithm being unable to make a clear choice.

The nature of an unclear choice was tested with the structure of WIN56291. The conformation of this structure was obtained from a crystallographic analysis of it, bound in another viral serotype, HRV-1A. In the light of the conformational dependence of the results reported above,

TABLE 3
SUPERIMPOSED RESULTS OF 3D (FROM 2D) STRUCTURES

	56590	51711	53438	56826	52035	52084	56442	52193	53338	58329	54721	54563	56291
56590	+10	+0	+10	+0	+0	+10	+10	+10	+10	+0	+0	+0	+0
51711	+0	+10	+0	+10	+10	+10	+9	+10	+10	+10	+10	+10	+10
53438	+10	+0	+10	+0	+0	+10	+10	+10	+0	+0	+0	+0	−1
56826	+0	+10	+0	+10	+10	+10	+10	+10	+10	+10	+10	+10	+10
52035	+0	+10	+0	+10	+10	+0	+8	+0	+10	+9	+10	+10	+10
52084	+10	+10	+10	+10	+0	+10	+0	+10	+0	+0	+0	+0	+0
56442	+10	+9	+10	+10	+8	+0	+10	+0	+10	+9	+10	+10	+10
52193	+10	+10	+10	+10	+0	+10	+0	+10	−1	+0	+0	+0	−2
53338	+10	+10	+0	+10	+10	+0	+10	−1	+10	+9	+10	+10	+10
58329	+0	+10	+0	+10	+9	+0	+9	+0	+9	+10	+10	+10	+7
54721	+0	+10	+0	+10	+10	+0	+10	+0	+10	+10	+10	+10	+10
54563	+0	+10	+0	+10	+10	+0	+10	+0	+10	+10	+10	+10	+10
56291	+0	+10	−1	+10	+10	+0	+10	−2	+10	+7	+10	+10	+10

it was of interest to see if the method would distinguish a structural conformation that was obtained in the same way from a different serotype. As the data in Table 1 show, this was the only structure where the orientational preference relative to the greatest majority of all the other structures was ambiguous (numbers close to ± 5). Clearly the method is a sensitive measure of the conformation chosen. This may be a serious limitation of the algorithm in some applications.

CONCLUSIONS

This study has shown that the method of simulated annealing of Dean et al. [8] does indeed do a consistent and reliable job of overlapping two molecules but that the resulting relative orientations are dependent on the starting conformations of the two molecules. A greater emphasis needs to be placed on methods of finding 'biorelevant' conformations in general. More refined calculations are needed in order to energetically distinguish between the overall change in free energy of a molecule in the series examined here as it is docked into the pocket in either of the two long axis orientations.

REFERENCES

1 Rossmann, M.G., Arnold, E., Erickson, J.W., Frankenberger, E.A., Griffith, J.P., Hecht, H.J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G., Rueckert, R.R., Sherry, B. and Vriend, G., Nature, 317(1985)145.

2 Smith, T.J., Kremer, M.J., Luo, M., Vriend, G., Arnold, E., Kamer, G., Rossmann, M.G., McKinlay, M.A., Diana, G.D. and Otto, M.J., Science, 283(1986)1286.

3 Staunton, D.E., Merluzzi, V.J., Rothlein, R., Barton, R., Marlin, S.D. and Springer, T.A., Cell, 56(1989)849.

4 Badger, J., Minor, I., Kremer, M.J., Oliveira, M.A., Smith, T.J., Griffith, J.P., Guerin, D.M.A., Krishnaswamy, S., Luo, M., Rossmann, M.G., McKinlay, M.A., Diana, G.D., Dutko, F.J., Fancher, M., Rueckert, R.R. and Heinz, B.A., Proc. Natl. Acad. Sci. U.S.A., 85(1988)3304.

5 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamling, R.C. and Kraut, J., J. Biol. Chem., 257(1982)13650.

6 Weber, P.C., Wendoloski, J.J., Pantoliano, M.W. and Salemme, F.R., J. Am. Chem. Soc., 114(1992)3197.

7 Pevear, D.C., Fancher, M.J., Felock, P.J., Rossmann, M.G., Miller, M.S., Diana, G., Treasurywala, A.M., McKinlay, M.A. and Dutko, F.J., J. Virol., 63 (1989) 2002 and references cited therein.

8 a. Barakat, M.T. and Dean, P.M., J. Comput.-Aided Mol. Design, 4(1990)295.
   b. Barakat, M.T. and Dean, P.M., J. Comput.-Aided Mol. Design, 4(1990)317.
   c. Barakat, M.T. and Dean, P.M., J. Comput.-Aided Mol. Design, 5(1991)107.
   d. Papadopoulis, M.C. and Dean, P.M., J. Comput.-Aided Mol. Design, 5(1991)119.

9 Smith, T.J., Kremer, M.J., Luo, M., Vriend, G., Arnold, E., Kamer, G., Rossmann, M.G., McKinlay, M.A., Diana, G.D. and Otto, M.J., Science, 283(1986)1286.

10 Clark, M., Cramer, R.D. and Opdenbosch, N.V., J. Comput. Chem., 10(1989)982.

11 Chemical Design Ltd., Oxford, U.K.