J Comput Aided Mol Des (2008) 22:105–109
DOI 10.1007/s10822-008-9173-z

Quantitative measurement of protease ligand conformation

Christopher J. R. Illingworth · Kevin E. B. Parkes · Christopher R. Snell · Christopher A. Reynolds

Received: 4 September 2007 / Accepted: 5 January 2008 / Published online: 19 January 2008

Abstract The tendency for protease ligands to bind in an extended conformation has been suggested as an important factor for the identification of compounds of medicinal importance. Here we present a novel graph-theoretical method giving a quantitative measure of ligand conformation, and through application of this method to a representative set of protease ligands in bound and unbound conformations, derive the result that protease ligands are more extended in conformation when in their bound state.

Keywords Conformation · Inhibitor · Ligand · Measurement · Protease

Introduction

As noted by Rich [8], protease ligands in their bound state tend to exhibit an extended beta-strand secondary structure in order to stabilize binding to the enzyme. This observation has been supported by examination of Ramachandran plots of peptidic protease ligands [9], and by superpositions of X-ray crystal structures, thus confirming the prevalence of extended conformation in most bound protease ligands [6, 10] and suggesting that protease ligand conformation is an important element in drug design [2]. One explanation is that, when in an extended conformation, more of the backbone of a peptide chain is open to interaction with the protease, increasing the likelihood of binding [10]. Previous work involved examination of protease ligands in their bound states alone, or used properties of the molecule based on eigenvalue decomposition or radius of gyration to compare conformations [1]. Here we present an alternative quantitative measure of ligand conformation, applicable to both peptidic and non-peptidic ligands, which allows for an objective comparison between ligands in their bound and unbound states. Relative to the eigenvalue and radius of gyration methods, our method has the advantage of being unaffected by the conformation of large side chains close to the centre of a ligand molecule. Applying this measure to a representative sample of protease ligands, we obtain a quantitative statistical confirmation of the prevalence of extended conformation in bound protease ligands.

Methods

A representative sample of protein–ligand systems was chosen from the MEROPS protease database [7], by choosing the highest resolution structure listed for each protease in the database. Proteases of all catalytic types were considered, with the exception of glutamic proteases, of which only one example was listed at the time of sampling. Systems were rejected if they had ligands that were very small, in which case the idea of conformation has little meaning in this context, or where the ligand had more than 12 residues. Inhibitors were then partitioned according to the catalytic type of the protease. This resulted in a set of 89 ligands, from the following pdb files: 1A16, 1A5I, 1AEC, 1AF0, 1AFQ, 1AO0, 1ARC, 1B12, 1B6A, 1BDA,

Electronic supplementary material The online version of this article (doi:10.1007/s10822-008-9173-z) contains supplementary material, which is available to authorized users.

C. J. R. Illingworth · C. A. Reynolds (✉)
Department of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK
e-mail: reync@essex.ac.uk

K. E. B. Parkes · C. R. Snell
Medivir UK Ltd., Chesterford Research Park, Little Chesterford, Essex CB10 1XL, UK

1BH6, 1BIL, 1BIO, 1BMQ, 1BQY, 1BRU, 1BXO, 1C1P, 1C24, 1CGH, 1CP3, 1CV8, 1CVR, 1CVZ, 1CZI, 1D4L, 1DMT, 1DY9, 1EAG, 1EKB, 1F2O, 1FH0, 1FJS, 1FN8, 1FT7, 1FXY, 1G9B, 1GA6, 1GEC, 1GKD, 1GMY, 1GTJ, 1GVU, 1H8L, 1HNE, 1HPG, 1HS6, 1HY7, 1I76, 1IAU, 1J2Q, 1KAP, 1KUG, 1LCP, 1LS5, 1LYB, 1M4H, 1ME4, 1MHW, 1MMQ, 1MU0, 1N1M, 1NJT, 1NQC, 1O86, 1ONX, 1OT5, 1P12, 1PFX, 1PJP, 1PWU, 1PYO, 1QDU, 1QJJ, 1QRP, 1QS8, 1S4V, 1SCN, 1SMR, 1UK4, 1WHT, 2RMP, 3APR, 3PRK, 4AIG, 5SGA, 6TMN, 830C, 966C.

For each ligand, the conformation was measured in two different ways. Firstly, a graph-theoretical algorithm was used to derive an absolute measure of the physical shape of the molecule. Each ligand was converted into a graph, with atoms represented by nodes, and bonds represented by edges between the atoms. Given any two atoms, labelled i and j, in the ligand, the distance in the graph $E(i, j)$ between the two atoms was defined as the number of edges in the shortest path in the graph between the atoms i and j, while the distance $D(i, j)$ was defined as the straight-line distance in three-dimensional space between the atoms i and j. Note that $E(i, j)$ is independent of ligand conformation, but $D(i, j)$ varies with conformation. Using the algorithm, heavy (non-hydrogen) atoms i and j were found in order to maximize the distance $E(i, j)$, and the path between i and j was found. Note that as $E(i, j)$ is the length of the shortest path between atoms, the possibility of going the wrong way around a ring is excluded. As many protease ligands are peptide-like in nature, this path can be thought of as being an approximation to the backbone of the ligand.

Supposing that the path that maximizes the distance $E(i, j)$ consists of the atoms $a_1, a_2, a_3, \ldots, a_{n-1}, a_n$, where $a_1 = i$ and $a_n = j$, the conformational score r was calculated by the formula

$$r = \frac{D(a_1, a_n)}{\sum_{k=1}^{n-1} D(a_k, a_{k+1})} \quad (1)$$

This gives the ratio between the straight line distance from atom i to atom j, and the distance in three-dimensional space along the bonds of the path from atom i to atom j. Hence the value r, which varies between 0 and 1, gives a measure of the conformation of the molecule. A value of 1 would indicate that the path from i to j along the bonds of the molecule was a perfectly straight line. A value of 0, while impossible for any ligand with more than one atom, would occur if the backbone of the molecule started and finished at exactly the same point in space.

In cases where there was more than one shortest path between the atoms i and j identified above, the values of r were calculated for each of these paths, and the conformational score was defined as the mean of these values.

In order to avoid potential errors in finding the backbone caused by large side chains at the end of the backbone, ligands from the systems examined were inspected by hand, and errors in the path were corrected as appropriate.

Secondly, the absolute conformational score for each ligand was compared to scores for a set of random conformations of the same ligand, in order to give a relative measure of conformation. For each ligand a stochastic conformational search was carried out using the program MOE (Chemical Computing Group, 2007) [3], in order to generate a representative set of unbound conformations of the ligand. The stochastic process generates random conformations of a ligand in vacuo, then minimizes the structure using the MMFF94x [4, 5] force field to find a low energy conformation. In each case an RMSD tolerance was chosen, and conformations of the ligand were generated. Each new conformation was compared to the conformations already in the set, and added to the set if it differed from the ligands in the set by at least the RMSD value. New conformations that were within the RMSD tolerance of at least one other conformation in the set were rejected, and the process was repeated until 20 successive new conformations of the ligand were rejected. Conformations with an internal energy 70 kcal mol^{-1} or more above the minimum energy of a conformation of that ligand were also rejected. The RMSD tolerance was set manually at a value between 0.1 and 3.8 Å, depending broadly on the size of the ligand, in order that the resultant set had between 100 and 1,000 different conformations of the ligand, giving a representative sample of the low energy conformations that could be adopted by the ligand. Where an RMSD tolerance of 0.1 Å led to a set with fewer than 100 ligands, the process was repeated, this time generating new conformations until 200 successive new conformations of the ligand were rejected. Details of the RMSD tolerances and the number of conformations in each ligand set are given in supporting information.

On visual inspection, this process generally did not result in the two ends of the ligand coming together to form a loop. Having generated a set of random conformations of each ligand, the binding conformation was compared and ranked against the conformation of molecules from the random set, using the graph-theoretical algorithm.

The set of random low-energy conformations gives an indication of ligand conformation in the absence of enzyme. Thus the relative measure gives an indication of the effect of the enzyme on ligand conformation. If, as has been suggested, proteases selectively bind the ligand in an extended conformation [2], it would be expected that ligands would be more extended in conformation when they are bound to an enzyme than when they are in an unbound state.

For each catalytic type of proteases, a set score, S, was generated according to the ranking of the bound ligand conformation compared to the random ligand conformations. Each ligand was assigned the value 1 if it was ranked in the top third of random conformations (i.e. the bound ligand was more extended in conformation than would be expected by chance), 2 if it was in the middle third of random conformations, or 3 if it was in the bottom third of random conformations (i.e. the bound ligand was less extended in conformation than would be expected by chance). The set score, S, was then defined as the sum of these values for all of the ligands in the set, a low score indicating generally more extended conformations, and a high score indicating generally less extended conformations. Statistical analysis was carried out by means of a trinomial distribution. Assuming that there are y ligands in a set, and that ligands are no more or less extended in conformation than would be expected by chance, the probability that S is less than or equal to some integer value x is given by the equation

$$P(S \leq x) = \sum_{y \leq \alpha+2\beta+3\gamma \leq x} \frac{(\alpha+\beta+\gamma)!}{\alpha!\beta!\gamma!}\left(\frac{1}{3}\right)^{\alpha+\beta+\gamma} \quad (2)$$

Here the dummy variables in the sum, α, β, and γ, represent the number of ligands in the set S obtaining the scores 1, 2, and 3, respectively. The probability, $P(S \leq x)$, gives the level of statistical significance with which the null hypothesis, that protease ligands are no more extended in conformation than would be expected by chance, can be rejected.

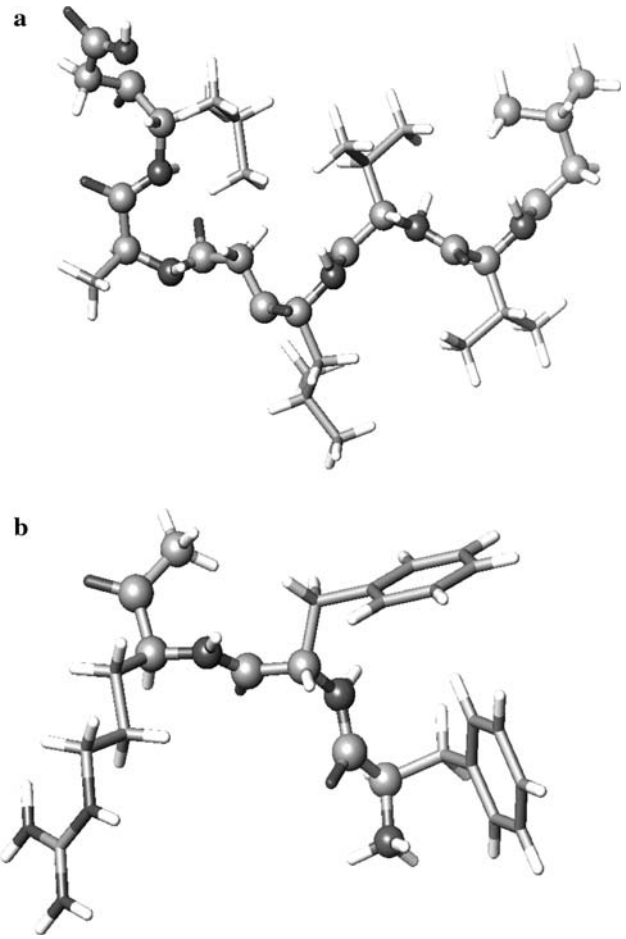

Fig. 1 Paths identified by the graph-theoretical algorithm for ligands in the systems (**a**) 2RMP and (**b**) 1CVR. In the former case, the algorithm has correctly identified the ligand backbone. In the latter case, side chain atoms have been excluded from the path

Results

Figure 1 shows the ligands for the systems 1CVR and 2RMP. The path identified by the algorithm is shown in ball and stick format, while other atoms are represented by sticks only. Note that for the ligand of 2RMP, the ligand backbone has been clearly identified by the algorithm. In the case of the system 1CVR side chains have been excluded from the path-finding algorithm in order to identify the backbone. Visual inspection of ligands led to the exclusion of side chains in 35 out of the 91 cases, namely 1AF0, 1ARC, 1BDA, 1BH6, 1BIL, 1BQY, 1CGH, 1CVR, 1CVZ, 1CZI, 1DMT, 1EKB, 1FH0, 1FN8, 1FXY, 1GMY, 1GTJ, 1GVU, 1HPG, 1HS6, 1IAU, 1J2Q, 1KAP, 1KUG, 1M4H, 1ME4, 1MHWb, 1OT5, 1PFX, 1PWU, 1PYO, 1S4V, 1SCN, 1SMR, 3APR.

Analysis of the bound protease ligands revealed a range of absolute conformational scores, with values of r ranging from 0.31 to 0.82 (Eq. 1). For individual ligands, the value of r was not always a strong indicator of the conformation of the ligand relative to other random conformations of the ligand. For example, the ligand of 1CVR had a value of r of 0.54, but was in the bottom third of the set of random conformations, while the ligand of 2RMP had a value of r of 0.45, but was in the top third of the random set. Figure 2 shows the distribution of the values of r for the 311 randomly generated conformations of the ligand 1A16. For this ligand, random values of r varied from 0.34 to 0.81, with the bound ligand, which is not included in the histogram data, receiving a score of 0.67. This gave the ligand a ranking of 144 out of 312 ligands, placing it in the middle third of the set of random conformations, and resulting in a set score S for this ligand of 2. Details of the conformational scores received by each of the ligands, including ranks and set scores, are given in supporting information.

When compared to a set of random conformations, the ligands exhibited a tendency towards being in a more extended conformation than would be expected by chance. Table 1 shows that for each type of proteases considered, a

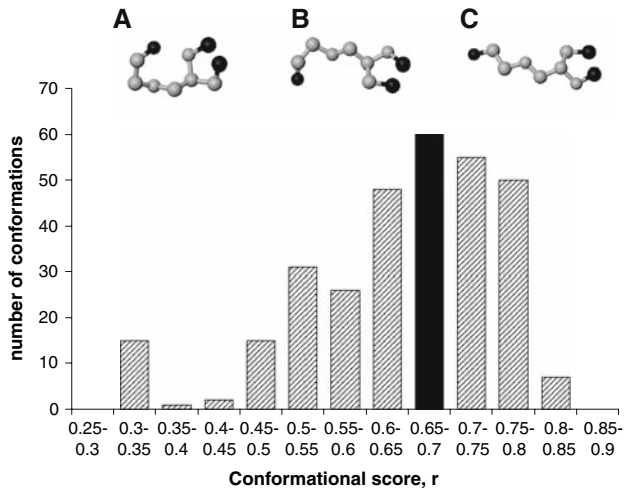

Fig. 2 Frequency of occurrence of conformational scores for 312 randomly generated conformations of the ligand from 1A16. The solid bar indicates the conformational score of the bound ligand. The molecular structures above the graph show the location of the atoms in the paths identified by the algorithm for (A) the conformation with minimal conformational score ($r = 0.34$), (B) the bound ligand conformation ($r = 0.67$), and (C) the conformation with maximal conformational score ($r = 0.81$). Note that the conformational score is calculated as the average of the two possible longest paths

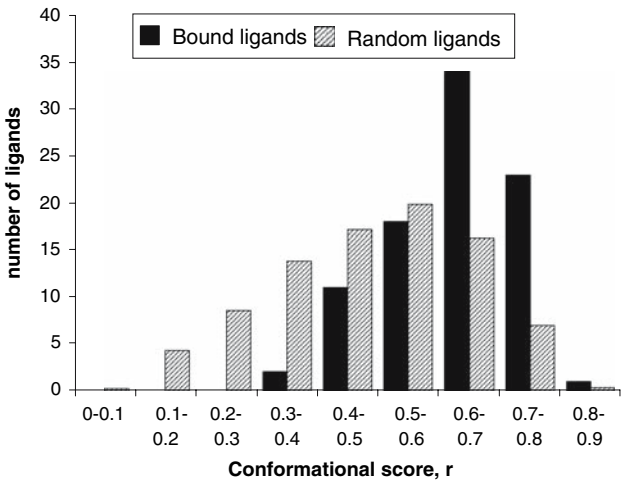

Fig. 3 Frequency of occurrence of conformational scores for the 89 ligands in the test set, and for the randomly generated conformations. The random conformations for each ligand are weighted to give a total of one ligand for each system. Bound ligands show a clear preference for more extended conformations

marked trend is observed for the bound ligands to be more extended in conformation than the random low energy unbound ligands. For each of the different catalytic types, the level of significance (>99.9%) was extremely high.

As an alternative method of comparing the conformations of ligands in bound and unbound conformations, the distributions of conformational scores for each were examined. For each ligand, the fraction of random conformations with each conformational score was calculated, and these fractions were summed over all ligands to give the representative distribution of random scores for the whole set. Figure 3 shows the distribution of conformational scores for the bound ligands compared to this representative random distribution. A clear preference for extended conformations in the set of bound ligands can be seen. Taken as a whole, this data gives objective statistical confirmation of the result [8, 10] that proteases selectively bind ligands in an extended conformation.

Discussion

We have presented here two measures for ligand conformation, the first an absolute measure, giving a description of the physical properties of the ligand, and the second a relative measure, evaluating whether a ligand is more extended in conformation than would be expected by chance.

The absolute measure for conformation described above differs from the more traditional approach of looking at dihedral angles, though a length of protein with an extended β-strand conformation would score very highly. This lack of dependence on dihedral angles bears some

Table 1 Analysis of the conformation of bound ligands from proteases of different catalytic types

Catalytic type	Number of ligands in set	Number of ligands with score 1, 2, and 3			Set score *S*	Significance of result (%)
		1	2	3		
Aspartic peptidases	14	12	2	0	16	99.997
Cysteine peptidases	15	11	3	1	20	99.906
Metallopeptidases	25	16	6	3	37	99.911
Mixed peptidases	24	21	3	0	27	99.999
Serine peptidases	11	9	1	1	14	99.901

A score of 1 indicates that the bound ligand is in a more extended conformation than would be expected by chance. A score of 3 indicates that the bound ligand is in a less extended conformation than would be expected by chance. A score of 2 indicates that the bound ligand is neither more or less extended in conformation than would be expected by chance

advantage in that the method is universally applicable to peptidic and non-peptidic ligands alike.

In the application of the absolute measure to evaluate ligand conformation, two challenges present themselves. In the first place, small changes in conformation at the end of ligands can have a large influence on the conformation score. For peptidic ligands where there is a large side chain towards the end of the backbone, this will by default be incorporated into the longest path found by the algorithm, though this can be corrected for manually. Secondly, as has been shown, a high or low absolute conformational score does not necessarily imply a high or low relative conformational score. Where the graph-theoretical method is used to give an indication of the conformation of a ligand, it is necessary to take into account the possible conformational space of that ligand.

Taken as a whole, however, the comparison of ligand conformation in bound and unbound states gives strong statistical support to the hypothesis of conformational selectivity in protease ligand binding. Protease ligands are more extended in conformation when they are in a bound state than would be expected when they are not bound to an enzyme.

References

1. Diller DJ, Merz KM Jr (2002) J Comput Aided Mol Des 16:105
2. Fairlie DP, Tyndall JD, Reid RC, Wong AK, Abbenante G, Scanlon MJ, March DR, Bergman DA, Chai CL, Burkett BA (2000) J Med Chem 43:1271
3. Ferguson DM, Raber DJ (1989) J Am Chem Soc 111:4371
4. Halgren TA (1996) J Comput Chem 17:490
5. Halgren TA, Nachbar RB (1996) J Comput Chem 17:587
6. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Structure 10:369
7. Rawlings ND, Morton FR, Barrett AJ (2006) Nucleic Acids Res 34:D270
8. Rich DH (1990) In: Hansch C, Sammes PG, Taylor JB (eds) Comprehensive medicinal chemistry, vol 2. Pergamon Press, Oxford, p 391
9. Tyndall JD, Fairlie DP (1999) J Mol Recognit 12:363
10. Tyndall JD, Nall T, Fairlie DP (2005) Chem Rev 105:973