Theor Chem Acc (2001) 106:128–136
DOI 10.1007/s002140000213

Theoretical
Chemistry Accounts

*Regular article*

# Analysis of the RGD sequence in protein structures: comparison to the conformations of the RGDW and $_{D}$RGDW peptides determined by molecular dynamics simulations

**Roland H. Stote**

Laboratoire de Chimie Biophysique, Université Louis Pasteur – ISIS, CNRS ESA-7006, 4 rue Blaise Pascal, 67000 Strasbourg, France

Received: 15 August 2000 / Accepted: 4 October 2000 / Published online: 21 March 2001


**Abstract.** Geometric properties of the RGD sequence in a data set of protein crystal and NMR structures deposited in the Protein Data Bank were examined to identify structural characteristics that are related to cell adhesion activity. Interatomic distances and dihedral angles are examined. These geometric measures are then used in an analysis of the conformations of the RGDW and $_{D}$RGDW peptides obtained from molecular dynamics simulations (Stote RH, et al. (2000) *J Phys Chem B* 104:1624). This analysis leads to the suggestion that differences in the accessible conformations contribute to the difference in biological activity between the RGDW and the $_{D}$RGDW peptides.

**Key words:** RGD – Conformation – Molecular dynamics – Cell adhesion

## 1 Introduction

The tripeptide sequence Arg-Gly-Asp (RGD) has been identified as a fundamental unit involved in recognition by a number of cell surface proteins [1] which play an important role in cell adhesion. NMR spectroscopy and X-ray crystallography have been used to the elucidate the structure of the RGD tripeptide in cell adhesion proteins [2–4], in RGD containing integrin inhibitors isolated from the venom of various vipers [5–8] and leeches [9], in virus coat proteins [10, 11], in small peptides containing the RGD sequence [12–16] and in small synthetic molecules containing the RGD moiety [17, 18]. There is, however, little information concerning the structure of the RGD receptor. In the absence of structural data for the receptor, insight into the relationship between structure and activity can be obtained by studying molecules that show a suitable affinity for the receptor. In the case of RGD-containing molecules, this has been done for small linear and cyclic peptides [14, 19–29] and small molecular analogs [30–35] by NMR and X-ray crystallography combined with activity assays. Conformationally constrained cyclic peptides are of particular interest since cyclization reduces the conformational freedom and provides an opportunity to stabilize a preferred conformation, thus reducing the entropy cost upon folding. In several cases, cyclic RGD peptides have shown higher activity than their linear counterparts. While such an analysis cannot provide conclusive results on the active conformation(s), it can suggest factors to consider in the design of potential therapeutics.

Another useful approach for obtaining insight into the conformational properties of RGD-containing peptides and proteins is by molecular dynamics simulations. Recently, molecular dynamics simulations of the biologically active peptides RGDW and $_{D}$RGDW have been reported [36]. Good agreement was found between the NMR experiments and the simulations. The NMR experiments suggested that the peptides are in a type II′ $\beta$ turn in solution, but that there are fast internal motions [16]. However, owing to the time scale of the internal motions, alternative conformations could not be determined. Molecular dynamics simulations were done to complement the NMR experiments. Analysis of the simulation trajectories indicated that a turn conformation is present in solution, consistent with the NMR experiments, but that an extended conformation is also present with a significant population in solution. Continuum dielectric (Poisson–Boltzmann) calculations suggested that the equilibrium constant of the turn conformer and extended conformer is near unity. The simulations also showed that the extended conformations are nearly indistinguishable from the $\beta$-turn conformation with respect to the experimental NMR results. Further analysis showed that the best fit to all the experimental data was from a mixture of the turn and the extended conformations.

Experimental studies suggest that cell adhesion activity is critically dependent on the conformation of the RGD sequence. Between RGDW and $_{D}$RGDW

Contribution to the Symposium Proceedings of Computational Biophysics 2000

peptides, an order of magnitude difference in activity has been measured; the latter is more active than the former. The cyclic peptide c(RGDW)2 was found to be biologically inactive. Although the difference in activity between $_D$RGDW and RGDW may be due, in part, to differences in the rate of biodegradation, it is of interest to examine the conformational differences between the two peptides in more detail. In the present work, we compare the conformations of the RGDW and $_D$RGDW peptides obtained from the molecular dynamics simulations to the conformations of the RGD sequence in proteins that are involved in cell adhesion. The geometric properties of the RGD sequence are first extracted from a data set of experimental protein structures deposited in the Protein Data Bank [37]. These geometric properties are then used to compare the conformations of the RGDW and $_D$RGDW peptides found in molecular dynamics simulations [36]. Some of these geometric measures have been previously used to characterize the structure–function relations in RGD-containing mimetics; they are used here to rationalize the difference in activity between the RGDW and $_D$RGDW peptides.

The methods are presented in Sect. 2; the results and discussion are presented in Sect. 3, followed by the conclusions.

## 2 Methods

A search of the Protein Data Bank [37] for proteins containing the sequence RGD using the web interface http://www.rcsb.org/pdb yielded 648 entries. For the analysis presented here, only those proteins where the RGD sequence is believed to exhibit cell adhesion functionality are used; this reduces the number of proteins used in the analysis to 26 structures (Table 1): 12 of these are X-ray structures and 14 structures were determined by NMR. To improve the statistics of the analysis, individual NMR structures were used rather than a single mean structure and in those proteins where identical multiple chains are present, all chains were used. While this introduces some overcounting, the conclusions remain unchanged. A total of 307 protein structures were used in the analysis and include, among others, proteins from the hepatitis C virus (NS3 protease), human collagenase and human rhinovirus, SH3 domain proteins and proteins extracted from the toxins of vipers, leeches and sea anemone. It is interesting to note that these distantly related proteins appear to be similar at the three-dimensional level with respect to the RGD sequence.

To characterize the conformations of the RGD sequence in the protein data set, the backbone $\phi$,$\psi$ dihedral angles of the Gly and Asp residues were calculated using the CHARMM program [38]; the resulting set of dihedral angles was then clustered using the ART-II′ algorithm implemented in the CHARMM program [38]. The ART-II′ algorithm is based on a self-organizing neural net [39–41]. The clustering algorithm describes the conformation of the peptide by a vector of $\boldsymbol{N}$ parameters; in the present study, the vector contains four parameters, which are the $\phi$ and $\psi$ backbone dihedral angles for the Gly and Asp residues. The clustering procedure produces a single, optimized partition of the data. The final partition of the data depends, in part, on the order in which the conformations are read, but previous work has shown that in the case of dihedral angles, the final cluster centers were not strongly affected by changes in the order of data input [36, 42].

To characterize common three-dimensional features of the RGD substructure in the protein data set, the pairwise atom distances between the C$\alpha$ and C$\beta$ atoms of the Arg and Asp residues, as well as the pairwise distance between the C$\zeta$ and the C$\gamma$ atoms of the Arg and Asp residues, respectively, were calculated. As

**Table 1.** Structures from the Protein Data Bank (*PDB*) used in the analysis

| PDB identification | Descriptor | Type[a] | Ref. |
|---|---|---|---|
| 1A1R | Viral protein | X 2 | [50] |
| 1AGI | Angiogenin | X | [51] |
| 1AHS | African horse sickness virus coat protein VP7 | X 3 | [52] |
| 1AYK | Collagenase | N 30 | [53] |
| 1AYM | Human rhinovirus 16 coat protein (subunit VP2) | X | [54] |
| 1AYN | Human rhinovirus 16 coat protein (subunit VP2) | X | [55] |
| 1AZE | GRB2 SOS | N 10 | [56] |
| 1BDS | BDS-I | N M | [57] |
| 1DAB | P.69 pertactin | X | [58] |
| 1DEC | Decorsin | N 25 | [9] |
| 1DKI | Pyrogenic exotoxin B zymogen mutant | X 4 | [69] |
| 1DRS | Dendroaspin (mambin S5C1/SH04) | N 39 | [59] |
| 1FNA | Fibronectin cell-adhesion module type III-10 | X | [60] |
| 1FNF | Fibronectin | X | [3] |
| 1FVL | Flavoridin | N 18 | [61] |
| 1JXP | NS3 serine protease NS4A | X 2 | [62] |
| 1KST | Kistrin | N 8 | [5] |
| 1MFN | Fibronectin | N 20 | [4] |
| 1PKT | Phosphatidylinositol 3-kinase (SH3 domain) | N 30 | [63] |
| 1TEN | Tenascin (third Fibronectin type III repeat) | X | [64] |
| 1TIV | HIV-1 transactivator protein; | N 10 | [65] |
| 1TTF | Fibronectin (tenth type III module) | N 36 | [66] |
| 2BDS | BDS-I (NMR 42 structures) | N 42 | [57] |
| 2ECH | Echistatin (NMR 8 structures) | N 8 | [67] |
| 2HWD | Rhinovirus 14 coat protein (subunit VP1, VP2) | X 2 | [68] |
| 2MFN | Fibronectin | N 10 | [4] |

[a] Specification of whether the structure was determined by X-ray crystallography (*X*) or by NMR (*N*). If by crystallography, the number of chains is specified if different than 1; if by NMR, the number of structures in the family is specified

a measure of the orientation of the Arg and Asp residues, the pseudodihedral angles defined by $\phi$[C$\beta$(***R***)-C$\alpha$(***R***)-C$\alpha$(D)-C$\beta$(D)] and $\phi$[C$\zeta$(***R***)-C$\alpha$(***R***)-C$\alpha$(D)-C$\gamma$(D)] were calculated. A $\phi$[C$\beta$(***R***)-C$\alpha$(***R***)-C$\alpha$(D)-C$\beta$(D)] pseudodihedral angle value close to zero indicates a parallel alignment, i.e., the sidechains are pointing in the same direction. A similar approach was taken in the study by Müller et al. [27], where the structure of 18 cyclic RGD peptides were determined by two-dimensional NMR spectroscopy and restrained molecular dynamics simulations; their activities were characterized by $IC_{50}$ values and were compared to the linear peptide GRGDS and to fibrinogen, vitronectin and fibronectin. From an analysis of interatomic distances and the pseudodihedral angle defined by $\phi$[C$\beta$(*R*)-C$\alpha$(*R*)-C$\alpha$(D)-C$\beta$(D)], it was shown that slight changes in conformation lead to significant differences in activity as measured by $IC_{50}$ values.

Molecular dynamics simulations of the RGDW and $_D$RGDW peptides were presented in earlier work [36, 43]; these simulations are used for the present analysis. The backbone dihedral angles from the trajectories were clustered using the ART-II′ algorithm under the same conditions as for the protein data set.

## 3 Results

The mean dihedral angles of the cluster centers resulting from the application of the ART-II′ clustering algorithm to the protein data set are given in Table 2. Both turn and extended conformations are represented by the mean cluster dihedral angles; for example, cluster 2 contains structures where the RGD sequence is in a type II′ $\beta$ turn and cluster 7 contains structures where the conformation of the RGD sequence is close to a type I $\beta$ turn. Protein structures that are members of these clusters include, among others, 1AGI, 1AYN, 1TEN, 1TIV, and 1DKI, 1DRS, respectively. Extended conformations are found in other protein structures, including 2BDS, 1FNA, 1PKT, 1FVI, 1DAB, 1AZE, 1TIV. In the family of 1KST structures, both extended and turn conformations are found; these are represented by clusters 2, 12 and 16. Using the mean dihedral angles of the cluster centers, several conformations of the RGD sequence were constructed showing turn and extended conformations (Fig. 1).

**Table 2.** Dihedral angle values for cluster centers of the protein data set

| Cluster number | Member population | Gly | | Asp | |
|---|---|---|---|---|---|
| | | $\phi$ | $\psi$ | $\phi$ | $\psi$ |
| 1 | 27 | 147.2 | −142.1 | −82.5 | −7.6 |
| 2 | 27 | 80.4 | −83.9 | −111.3 | −28.4 |
| 3 | 60 | 64.5 | −133 | −93.5 | 102.2 |
| 4 | 11 | 102.3 | −24.5 | −59.1 | 133.3 |
| 5 | 18 | −119.6 | −150.4 | −97.4 | 28.6 |
| 6 | 1 | −61.9 | 176.4 | 149.7 | −160.6 |
| 7 | 39 | −87 | −50.2 | −114.9 | 3.2 |
| 8 | 13 | −103.8 | 34.2 | 60 | 45.6 |
| 9 | 1 | −102.2 | −156.3 | 55.4 | 47.5 |
| 10 | 75 | −164.5 | 172.7 | −87.1 | 141.3 |
| 11 | 7 | 149.9 | 142.1 | 65.4 | 18.6 |
| 12 | 3 | −92.4 | 85.2 | −101.9 | 63 |
| 13 | 3 | −84.1 | 89.5 | −176.2 | −80.5 |
| 14 | 5 | −150.3 | −76.5 | −95.9 | 137 |
| 15 | 22 | 143.4 | 4 | −151.8 | −51.7 |
| 16 | 1 | 76.9 | 104.9 | −77.2 | 101.2 |

The results from the cluster analysis of the molecular dynamics simulations of the RGDW and $_D$RGDW peptides are given in Table 3. Note that in Ref. [36], the cluster analysis used both backbone and sidechain dihedral angles, whereas in the present analysis, only backbone dihedral angles are used. Both turn and extended structures are present for both the RGDW and $_D$RGDW peptides. In the present cluster analysis, turn conformations for the RGDW peptide are given by cluster centers 1 and 5, while the $_D$RGDW peptide tends to be in more distorted turn and extended conformations. A number of the cluster centers obtained from the analysis of the trajectories are similar to the cluster centers obtained from the analysis of the protein data set. Backbone conformations having dihedral angle values given by the cluster centers were constructed for all cluster centers and the root-mean-square (rms) coordinate deviations were calculated between the conformations derived from the protein data set and the conformations

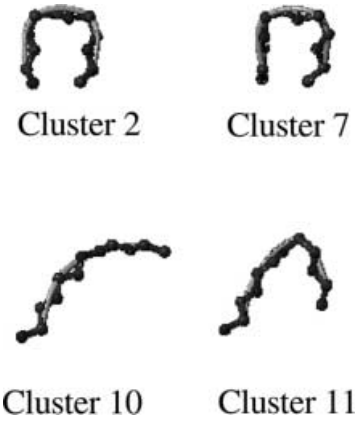


**Fig. 1.** Schematic drawing of backbone conformations of the RGD sequence derived from the cluster analysis of the protein data set, see Table 2. Cluster 2 is a type II′ $\beta$ turn, cluster 7 is a type I $\beta$ turn; clusters 10 and 11 are examples of extended and distorted turn structures, respectively

**Table 3.** Dihedral angle values for cluster centers from the molecular dynamics simulation of the RGDW and $_D$RGDW peptides

| Cluster number | Member population | Gly | | Asp | |
|---|---|---|---|---|---|
| | | $\phi$ | $\psi$ | $\phi$ | $\psi$ |
| RGDW | | | | | |
| 1 | 1322 | 86.0 | −91.3 | −84.4 | −46.0 |
| 2 | 1768 | −161.7 | −168.9 | −100.2 | −51.0 |
| 3 | 811 | −150.1 | −98.2 | −105.7 | −80.2 |
| 4 | 1851 | −165.1 | −80.9 | −118.0 | 171.4 |
| 5 | 973 | 92.9 | −64.8 | −106.1 | −57.3 |
| $_D$RGDW | | | | | |
| 1 | 972 | 93.1 | −149.0 | −85.7 | 174.0 |
| 2 | 1070 | 136.3 | 160.9 | −89.5 | −71.9 |
| 3 | 995 | −96.9 | −172.0 | −89.6 | −88.2 |
| 4 | 1971 | −135.6 | −96.6 | −90.5 | 136.7 |
| 5 | 1165 | −162.2 | −65.9 | −95.5 | −169.9 |
| 6 | 327 | 153.1 | 41.9 | −114.5 | −143.8 |

derived from the analysis of the molecular dynamics trajectories. rms differences less than 0.3 Å are observed in numerous cases; for example, between RGDW clusters 1 and 5 and the protein data set clusters 2 and 7. For the $_{D}$RGDW case, the rms differences between $_{D}$RGDW clusters 4 and 5 and the protein data set cluster 14 are also less than 0.3 Å. For an rms difference between 0.3 and 1.0 Å, there are many similar cluster pairs; the rms differences between cluster pairs are shown in Fig. 2. This analysis indicates that the simulations sample conformations of the RGDW peptides that are physically realistic and similar to the conformations of the RGD sequence found in biologically active proteins.

The interatomic distances, $D$[C$\alpha$($R$)-C$\alpha$(D)], $D$[C$\beta$($R$)-C$\beta$(D)] and $D$[C$\zeta$($R$)-C$\gamma$(D)] and the pseudodihedral angles $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] and $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)] were calculated for the proteins in the data set, from the molecular dynamics simulations [36] and from the 15252 RGDW structures determined in the adaptive umbrella sampling simulations of Bartels et al. [43]. The interatomic distance between the charges of the two sidechains, $D$[C$\zeta$($R$)-C$\gamma$(D)], and the pseudodihedral angle, $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)], are calculated as a measure of the distance and the relative orientation of the charged termini of the two sidechains, respectively.

The interatomic distances calculated from the protein data set and from the RGDW and $_{D}$RGDW simulations are shown in Fig. 3. All data sets give about the same results for $D$[C$\alpha$($R$)-C$\alpha$(D)], with the values ranging between 5 and 7.5 Å, and for $D$[C$\beta$($R$)-C$\beta$(D)], the values range between 5.5 and 9 Å. The C$\zeta$($R$)-C$\gamma$(D) distances, $D$[C$\zeta$($R$)-C$\gamma$(D)], calculated from the structures in the protein data set are more widely distributed and are larger than from the simulations (between 7.8 and 11.5 Å); this is probably due to a lack of nuclear Overhauser effect for the termini of the sidechains. Aside from this, little difference between the distances calculated from the simulations and from the proteins was found; the distances $D$[C$\alpha$($R$)-C$\alpha$(D)], and $D$[C$\beta$($R$)-C$\beta$(D)], fall into the ranges defined by the results of Müller et al. [27].

The relative orientation of the Arg and Asp sidechains is measured by the pseudodihedral angle, $\phi$ [C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)], as described previously. The distributions of $\phi$[C$\beta$(R)-C$\alpha$(R)-C$\alpha$(D)-C$\beta$(D)] are shown in Fig. 4, where the overall distribution and the distributions from the individual simulations are given. In the distribution of pseudodihedral angles calculated from the protein data set, there are two predominant peaks. One relatively broad peak covers the range between $-100°$ and $+10°$. A second peak goes from about 30° to $+100°$. This is consistent with the distribution found by Müller et al. [27], where the most active small molecules have a geometry that closely resembles that of the RGD sequence in biologically active proteins. The values of $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] from the RGDW simulations (Fig. 4B) range from $-60°$ to 180° with a dominant peak around 0°. Of the four RGDW simulations, one shows no population in the "active" region around 0.0° and a large distribution between 60° and 180°; the other three RGDW simulations sample conformations in the "active" region. In the umbrella sampling simulations of Bartels et al. [43] (Fig. 4C), the RGDW peptide tends to populate regions outside the active region, although a small population in the active region is found. Examination of the relative probabilities from this simulation indicates that the structures in the active region are less probable than those outside the active region. In fact, the 39 most probable structures from this set have pseudodihedral angles around $-100°$. This suggests that the RGDW simulations presented in Ref. [36] may access conformations of somewhat lower probability but still of physiological significance. In the

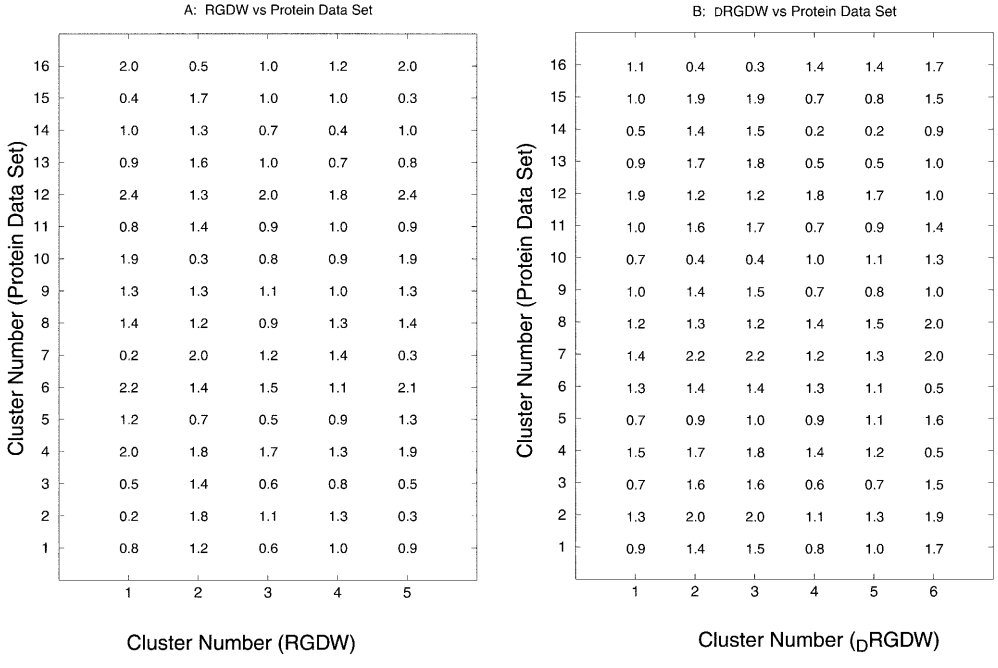


**Fig. 2.** Root-mean-square coordinate difference between the backbone conformations constructed using the mean cluster dihedral angles of the protein data set and the mean cluster dihedral angles of the molecular dynamics conformations of the RGDW (**A**) and $_{D}$RGDW (**B**) peptides

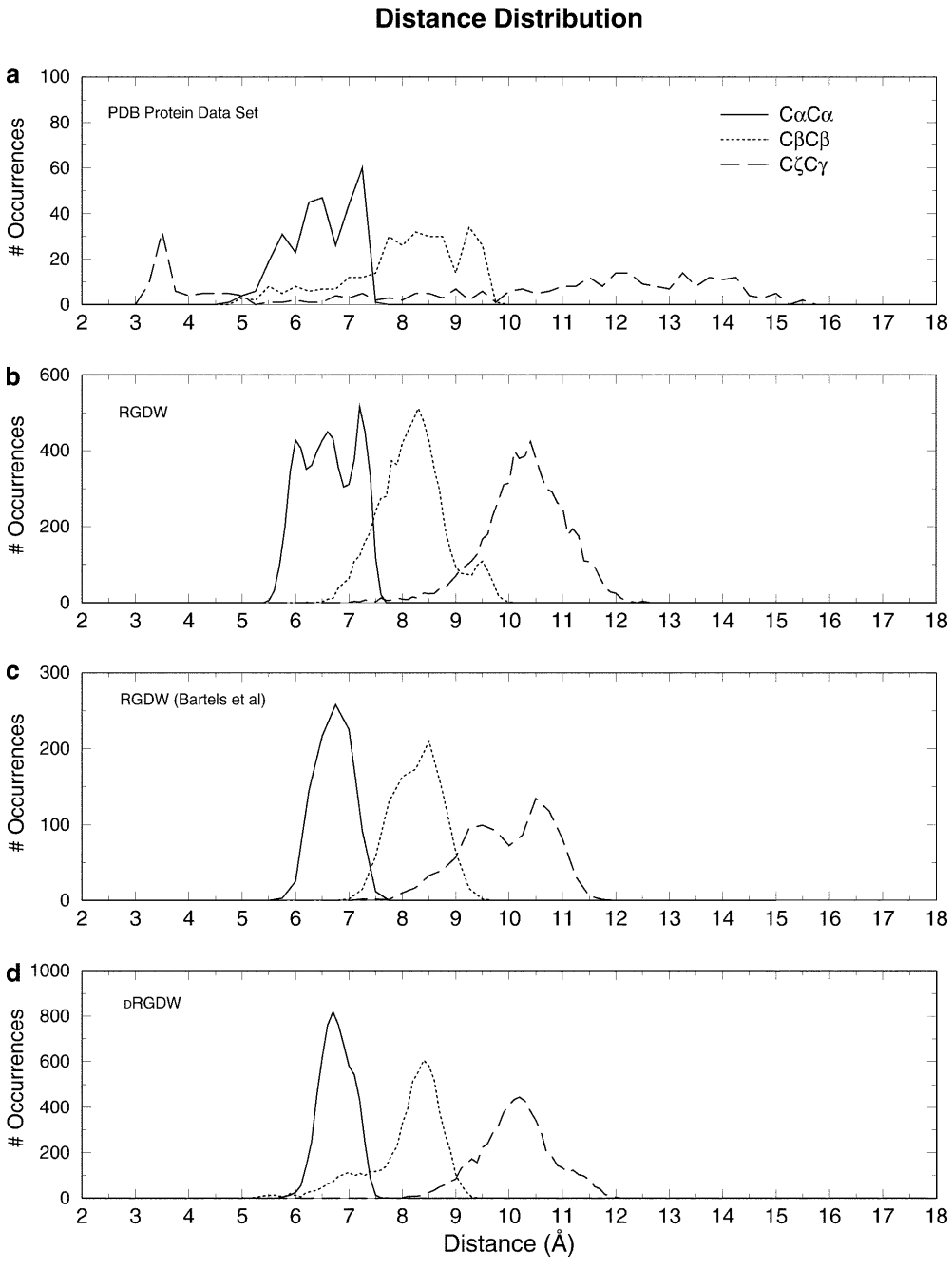


**Fig. 3.** Distribution of interatomic distances from RGD conformations. The *solid line* is the distribution for the distance Arg:Cα-Asp:Cα the *dotted line* is for the distance Arg:C*β*-Asp:C*β* and the *dashed line* is the distribution for the distance Arg:Cζ-Asp:C*γ* **a** from the protein data set structures (see text), **b** from the RGDW simulations from Stote et al. [36], **c** from the adaptive umbrella sampling simulations of RGDW from Bartel et al. [43] and **d** from $_{D}$RGDW simulations of Stote et al. [36]

$_{D}$RGDW simulations (Fig. 4D), there is a significant distribution of $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] centered at about −30° that drops to zero at the boundaries of the "active" region. An additional population around 160° and −170° is found. The population in the active region is due mainly to one simulation ($_{D}$RGDW_3 in Ref. [36]); the other two simulations show no population in this region. One difference between the RGDW and the $_{D}$RGDW results is that just outside the active region the population drops to zero for $_{D}$RGDW, while it does not for RGDW. The narrower distribution for $_{D}$RGDW suggests that, as for constrained cyclic peptides, the entropy cost for assuming a conformation in the "active" region is lower for the $_{D}$RGDW peptide than for the RGDW peptide. The constrained conformation is probably due to a hydrogen-bond interaction between the carboxy terminal of the Asp sidechain and the amide hydrogen on the Trp residue as discussed in Ref. [36]. This interaction was not present in the simulations of the RGDW peptide.

The relative orientation of the charged ends of the Arg and Asp sidechains is measured here by the pseudodihedral angle $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)]. Pseudodihedral angle maps of $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)] versus $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] are shown in Fig. 5. The results from the analysis of the protein data set (Fig. 5A), from the RGDW simulations (Fig. 5B), from the adaptive umbrella sampling simulations of Bartels et al. (Fig. 5C) and from the $_{D}$RGDW simulations (Fig. 5D) are shown. The proteins in the data set

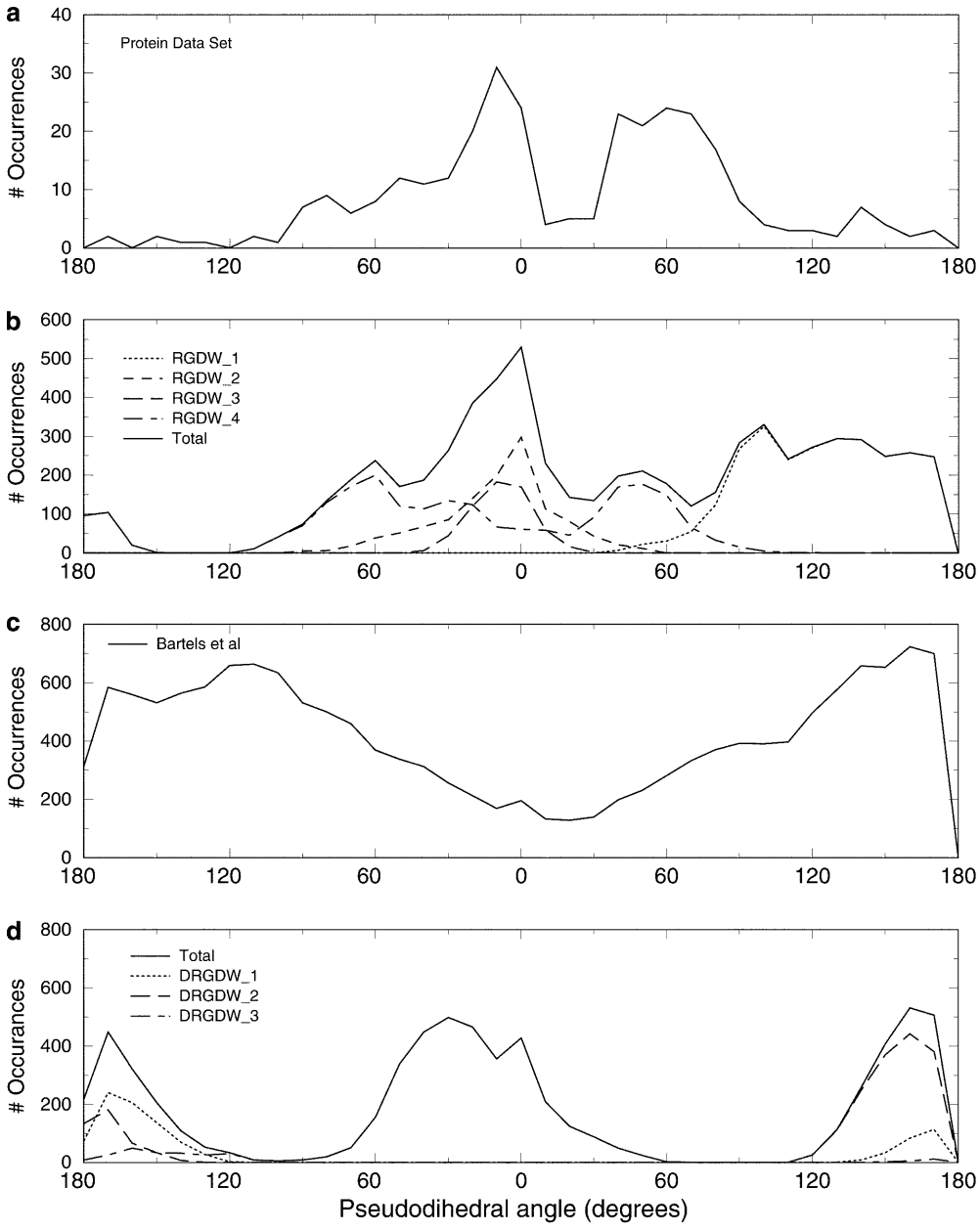


**Fig. 4.** Distribution of pseudodihedral angle $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] from RGD conformations **a** from the protein data set structures (see text), **b** from the RGDW simulations from Stote et al. [36], **c** from the adaptive umbrella sampling simulations of RGDW from Bartels et al. [43] and **d** from $_{D}$RGDW simulations of Stote et al. [36]

populate a region of dihedral angle space bounded by −120° and 100° for $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] and between −150° and 90° for $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)]. Both the RGDW and the $_{D}$RGDW simulations have conformations that are in this region of conformation space, although the distribution for $_{D}$RGDW is more localized than that for RGDW.

The geometric properties for interatomic distances and pseudodihedral angles were calculated for five proteins in which the RGD sequence is not involved in cell adhesion: these are the biotin carboxylase subunit of acetyl-CoA carboxylase (1BNC) [44], nucleoside diphosphate kinase (1NSQ) [45], α-lytic protease complex (1P01) [46] and d-xylose isomerase (1XIB [47] and 1XLA [48]). For these proteins, the C$\alpha$-C$\alpha$ distances ranged from 5.3 to 7.3 Å, the C$\beta$-C$\beta$ distances ranged from 6.0 to 9.6 Å and the C$\zeta$-C$\gamma$ distances ranged from 9.4 to 14.3 Å. These distances fall in the same range as for the cell adhesion active proteins. The values of the pseudodihedral angle $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] for 1XIB, 1XLA and 1BNC were in the range −64.5° to −79.4°, which is in the wings of the active region distribution. The values of $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] fall outside the active range defined by the cell adhesion proteins for 1NSQ (133.75°) and for 1P01 (144.9°). For the pseudodihedral angle $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)], the values covered the range from −16.0° to −165.6°, which are also at the extremes of the active region.

These results suggest that the activity of RGD depends not only on the pseudodihedral angle $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)], which defines the relative orientation of the Arg and Asp sidechains, but also on the relative orientation of the two charged termini defined here by the pseudodihedral angle $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)]. The angle $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] coarsely defines the regions of conformational

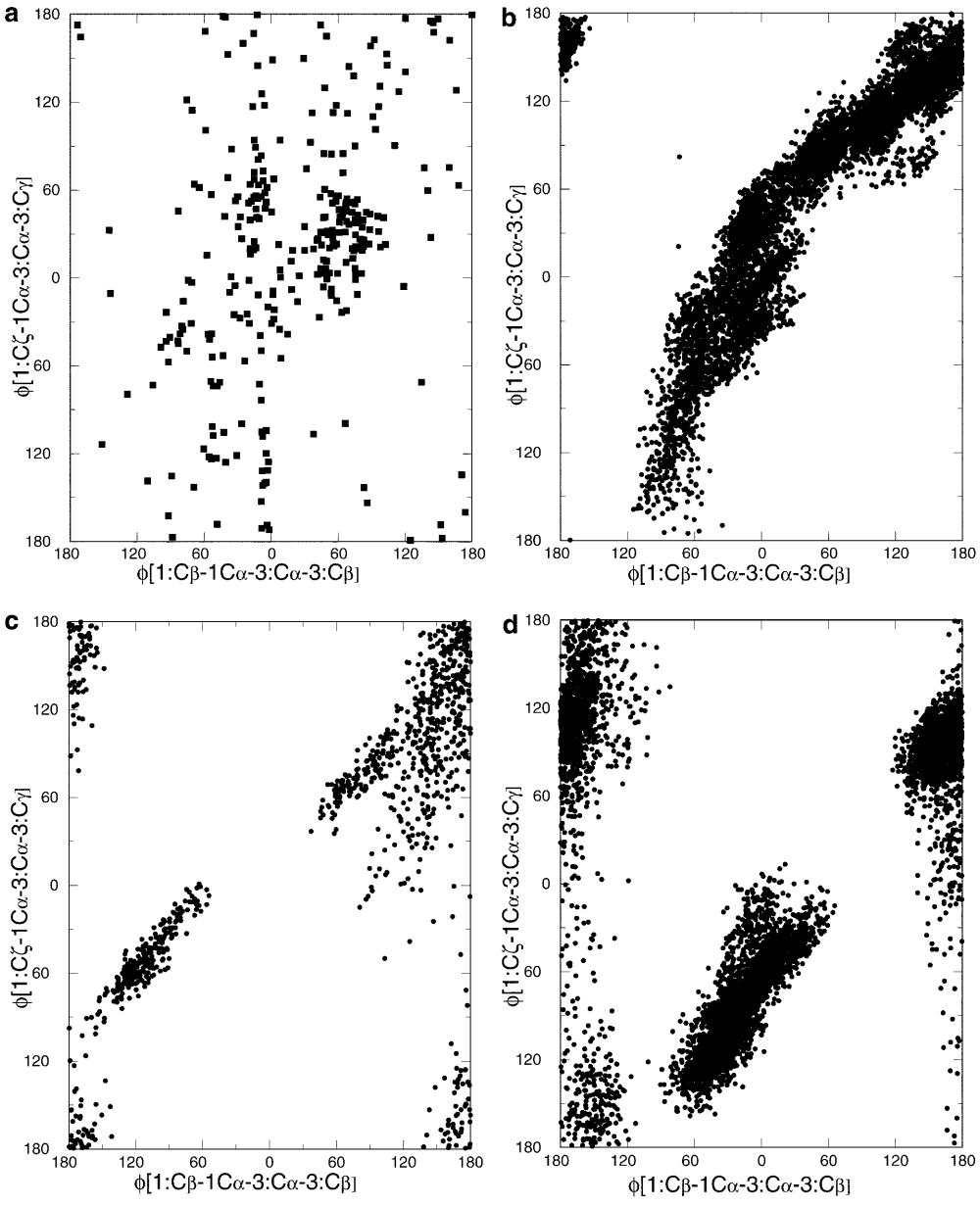


**Fig. 5.** Correlation diagram of the pseudodihedral angle $\phi$[C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] versus $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)] calculated from RGD conformations **a** from the protein data set structures (see text), **b** from the RGDW simulations from Stote et al. [36], **c** from the adaptive umbrella sampling simulations of RGDW from Bartel et al. [43] and **d** from $_{D}$RGDW simulations of Stote et al. [36]

space available to the sidechains, while the pseudodihedral angle $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)] may be important for optimizing the interactions. As mentioned previously, the narrower distribution of dihedral angles for $_{D}$RGDW in the active region of the conformational space may reduce the entropy cost for assuming an active conformation. This is the basis for the design of constrained cyclic peptides and is likely to contribute to the increased activity observed for the $_{D}$RGDW peptide relative to the RGDW peptide. The entropy difference between the pseudodihedral angle distributions for the two peptides can be estimated from a quasiharmonic model [49] by the equation

$$\Delta S = S_b - S_a = \frac{k_b}{2} \ln \frac{\sigma_b}{\sigma_a} \quad ,$$

where $\sigma$ is the determinant of the covariance matrix $\sigma$ with elements $\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle$ and $q_i$ and $q_j$ are the individual pseudodihedral angles and the angled brackets represent averages over the simulation. The calculation of the entropy difference between the two distributions was limited to the active region for $\phi$ [C$\beta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\beta$(D)] ($\pm 120°$) and without bounds on $\phi$[C$\zeta$($R$)-C$\alpha$($R$)-C$\alpha$(D)-C$\gamma$(D)]. The entropy for the $_{D}$RGDW distribution was 1.18 cal/kmol less than that for RGDW, giving a $\Delta G$ of 0.35 kcal/mol at 300 K.

## 4 Conclusions

The geometry of RGD sequences in proteins involved in cell adhesion was examined for 307 protein structures from 26 entries in the Protein Data Bank and for five protein structures that are not involved in cell adhesion. The results show that both turn and extended structures of the RGD sequence are present in these proteins.

Geometric properties, which include pairwise distances between the Cα and Cβ atoms in the Arg residue and the corresponding atoms in the Asp residues and pseudodihedral angles between these two residues, were calculated. These results were compared to the results obtained from an analysis of the conformations of the RGDW and $_{D}$RGDW peptides obtained from molecular dynamics simulations. RGDW and $_{D}$RGDW are both active in the inhibition of platelet aggregation; $_{D}$RGDW is an order of magnitude more active than RGDW. Comparison between the conformations sampled in molecular dynamics simulations and available protein structure data, including the X-ray crystal and NMR solution structures of active RGD-containing proteins and 18 small cyclic RGD peptides [28], gives further insight into the dependence of biological activity on conformation. Two aspects of sidechain orientation are likely to be important for biological activity. The first is that the Arg and Asp sidechains are on the same side of the peptide in an almost parallel alignment. This was identified as a factor in studies of Müller et al. [27]. The second aspect is the orientation of the two charged termini of the Arg and Asp sidechains. Optimal regions of conformation space are defined in the present work. For the cell adhesion inactive proteins, the values of the pseudodihedral angle measured tended to be at the extremities or outside of the active region. The difference in activity between the RGDW and $_{D}$RGDW peptides may be due, in part, to differences in configurational entropy within this active region of the conformation space. The narrower distribution of dihedral angles involving the residues Arg and Asp in the active region suggests that the entropy loss upon folding to an ''active'' conformation would be less for the $_{D}$RGDW peptide than for the RGDW peptide. This, in turn, may lead to a more active peptide. Information derived from this study should be of use in the design of potential therapeutic drugs for thrombosis and other diseases where much of the rational drug design strategy has been based on assumed bioactive conformations of the RGD sequence.

*Acknowledgements*. This work was supported, in part, by the Fondation pour la Recherche Médicale (France), the European Economic Community (Human Capital and Mobility) and the CNRS, France. The author would like to thank Annick P. Dejaegere and Martin Karplus for their helpful comments and discussion.

## References

1. Ruoslahti E, Pierschbacher MD (1986) Cell 44: 517
2. Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T (1983) J Mol Biol 170: 175
3. Leahy DJ, Ikramuddin A, Erickson HP (1996) Cell 84: 155
4. Copie V, Tomita Y, Akiyama SK, Aota S-I, Yamada KM, Venable RM, Pastor RW, Krueger S, Torchia DA (1998) J Mol Biol 277: 663
5. Adler M, Lazarus RA, Dennis MS, Wagner G (1991) Science 253: 445
6. Saudek V, Atkinson RA, Pelton JT (1991) Biochemistry 30: 7369
7. Cooke AM, Carter BG, Martin DMA, Murray-Rust P, Weir MP (1991) Eur J Biochem 202: 323
8. Gould RM, Polokoft MA, Friedman PA, Huang TF, Holt JT, Cook JJ, Niewiarowski S (1990) Proc Soc Exp Biol Med 195: 168
9. Krezel AM, Wagner G, Seymour-Ulmer J, Lazarus RA (1994) Science 264: 1944
10. Logan D, Abu-Ghazaleh R, Blakemore W, Curry S, Jackson T, King A, Lea S, Lewis R, Newman J, Parry N, Rowland SD, Stuart D, Fry E (1993) Nature 362: 566
11. Grimes J, Basak AK, Roy P, Stuart D (1995) Nature 373: 167
12. Reed J, Hull WE, von der Lieth C-W, Kübler D, Suhai S, Kinzel V (1988) Eur J Biochem 178: 141
13. Genest M, Marion D, Ptak M (1989) In: Aubry A, Marraud M, Vitoux B (eds) Second forum on peptides, vol 174. John Libbey Eurotext, p 415
14. Williamson MP, Davies JS, Thomas WA (1991) J Chem Soc Perkin Trans 2 601
15. Mickos H, Bahr J, Lüning B (1990) Acta Chim Scand 44: 161
16. Kieffer B, Mer G, Mann A, Lefèvre JF (1994) Int J Pept Protein Res 44: 70
17. Cachau RE, Serpersu EH, Mildvan AS, August JT, Amzel LM, (1989) J Mol Recognit 2: 179
18. Johnson CWJ, Pagano TG, Basson CT, Madri JA, Gooley P, Armitage IM (1993) Biochemistry 32: 268
19. Gurrath M, Müller G, Kessler H, Aumailley M, Timpl R (1992) Eur J Biochem 210: 911
20. McDowell RS, Gadek TR (1992) J Am Chem Soc 114: 9245
21. Kopple KD, Baures PW, Bean JW, D'Ambrosio C, Hughes JL, Peoshoff CE, Eggleston DS (1992) J Am Chem Soc 114: 9615
22. Davies JS, Enjalbal C, Wise CJ, Webb SE, Jones GE (1994) J Chem Soc Perkin Trans 1 2011
23. Bach AC II, Eyermann CJ, Gross JD, Bower MJ, Harlow RL, Weber PC, DeGrado WF (1994) J Am Chem Soc 116: 3207
24. Jackson S, DeGrado W, Dwivedi A, Parthasarathy A, Higley A, Krywko J, Rockwell A, Markwalder J, Wells G, Wexler R, Mousa S, Harlow R (1994) J Am Chem Soc 116: 3220
25. McDowell RS, Gadek TR, Barker PL, Burdick DJ, Chan KS, Quan CL, Skelton N, Struble M, Thorsett ED, Tischler M, Tom JYK, Webb TR, Burnier JP (1994) J Am Chem Soc 116: 5069
26. Geyer A, Müller G, Kessler H (1994) J Am Chem Soc 116: 7735
27. Müller G, Gurrath M, Kessler H (1994) J Comput Aided Mol Des 8: 709
28. Haubner R, Schmitt W, Hölzemann G, Goodman SL, Jonczyk A, Kessler H (1996) J Am Chem Soc 118: 7881
29. Wermuth J, Goodman SL, Jonczyk A, Kessler H (1997) J Am Chem Soc 119: 1328
30. Ku TW, Ali FE, Barton LS, Bean JW, Bondinell WE, Burgess JL, Callahan JF, Calvo RR, Chen L, Eggleston DS, Gleason JG, Huffman WF, Hwang SM, Jakas DR, Karash CB, Keenan RM, Kopple KD, Miller WH, Newlander KA, Nichols A, Parker MF, Peishoff CE, Samanen JM, Uzinskas I, Venslavsky JW (1993) J Am Chem Soc 115: 8861
31. Greenspoon N, Hershkoviz R, Alon R, Varon D, Shenkman B, Marx G, Federman S, Kapustina G, Lider O (1993) Biochemistry 32: 1001
32. McDowell RS, Blackburn BK, Gadek TR, McGee LR, Rawson T, Reynolds ME, Robarge KD, Somers TC, Thorsett ED, Tischler M, Webb RR II, Venuti MC (1994) J Am Chem Soc 116: 5077
33. Blackburn BK, Lee A, Baier M, Kohl B, Olivero AG, Matamoros R, Robarge KD, McDowell RS (1997) J Med Chem 40: 717
34. Fisher MJ, Gunn B, Harms CS, Kline AD, Mullaney JT, Nunes A, Scarborough RM, Arfsten AE, Skelton MA, Um SL, Utterback BG, Jakubowski JA (1997) J Med Chem 40: 2085
35. Sall DJ, Arfsten AE, Bastian JA, Denney ML, Harms CS, McCowan JR, Morin JM Jr., Rose JW, Scarborough RM, Smyth MS, Um SL, Utterback BG, Vasileff RT, Wikel JH, Wyss VL, Jakubowski JA (1997) J Med Chem 40: 2843
36. Stote RH, Dejaegere AP, Lefèvre J-F, Karplus M (2000) J Phys Chem B 104: 1624

37. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28: 235
38. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan SMK (1983) J Comput Chem 4: 187
39. Carpenter GA, Grossberg S (1987) Appl Opt 26: 4919
40. Pao Y-H (1989) Adaptive pattern recongition and neural networks. Addison-Wesley, New York
41. Karpen ME, Tobias DJ, Brooks CL III (1993) Biochemistry 32: 412
42. Tobias DJ, Mertz JE, Brooks CL III (1991) Biochemistry 30: 6054
43. Bartels C, Stote RH, Karplus M (1998) J Mol Biol 284: 1641
44. Waldrop GL, Rayment I, Holden HM (1994) Biochemistry 33: 10249
45. Morera S, Chiadmi M, Lebras G, Lascu I, Janin J (1995) Biochemistry 34: 11062
46. Bone R, Shenvi AB, Kettner CA, Agard DA (1987) Biochemistry 26: 7609
47. Carrell HL, Hoier H, Glusker JP (1989) Proc Natl Acad Sci USA 86: 4440
48. Collyer CA, Henrick K, Blow DM (1990) J Mol Biol 212: 211
49. Karplus M, Kushick JN (1981) Macromolecules 14: 325
50. Kim JL, Morgenstern KA, Lin C, Fox T, Dwyer MD, Landro JA, Chambers SP, Markland W, Lepre CA, O'Malley ET, Harbeson SL, Rice CM, Murcko MA, Caron PR, Thomson JA (1996) Cell 87: 343
51. Acharya KR, Shapiro R, Riordan JF, Vallee BL (1995) Proc Natl Acad Sci USA 92: 2949
52. Basak AK, Gouet P, Grimes J, Roy P, Stuart D (1996) J Virol 70: 3797
53. Moy FJ, Chanda PK, Cosmi S, Pisano MR, Urbano C, Wilhelm J, Powers R (1998) Biochemistry 37: 1495
54. Hadfield AT, Lee WM, Zhao R, Oliveira MA, Minor I, Rueckert RR, Rossmann MG (1997) Structure 5: 427
55. Oliveira MA, Zhao R, Lee WM, Kremer MJ, Minor I, Rueckert RR, Diana GD, Pevear DC, Dutko FJ, McKinlay MA (1993) Structure 1: 51
56. Vidal M, Goudreau N, Cornille F, Cussac D, Gincel E, Garbay C (1999) J Mol Biol 290: 717
57. Driscoll PC, Gronenborn AM, Beress L, Clore GM (1989) Biochemistry 28: 2188
58. Emsley P, Charles IG, Fairweather NF, Isaacs NW (1996) Nature 381: 90
59. Sutcliffe MJ, Jaseja M, Hyde EI, Lu X, Williams JA (1994) Nat Struct Biol 1: 802
60. Dickinson CD, Veerapandian B, Dai X-P, Hamlin RC, Xuong N-H, Ruoslahti E, Ely KR (1994) J Mol Biol 236: 1079
61. Senn H, Klaus W (1993) J Mol Biol 232: 907
62. Yan Y, Li Y, Munshi S, Sardana V, Cole JL, Sardana M, Steinkuehler C, Tomei L, De Francesco R, Kuo LC, Chen Z (1998) Protein Sci 7: 837
63. Koyama S, Yu H, Dalgarno DC, Shin TB, Zydowsky LD, Schreiber SL (1993) Cell 72: 945
64. Leahy DJ, Hendrickson WA, Aukhil I, Erickson HP (1992) Science 258: 987
65. Bayer P, Kraft M, Ejchart A, Westendorp M, Frank R, Rosch P (1995) J Mol Biol 247: 529
66. Main AL, Harvey TS, Baron M, Boyd J, Campbell ID (1992) Cell 71: 671
67. Atkinson RA, Saudek V, Pelton JT (1994) Int J Pept Protein Res 43: 563
68. Kim KH, Willingmann P, Gong ZX, Kremer MJ, Chapman MS, Minor I, Oliveira MA, Rossmann MG, Andries K, Diana GD (1993) J Mol Biol 230: 206
69. Kagawa TF, Cooney JC, Baker HM, McSweeney S, Liu M, Gubba S, Musser JM, Baker EN (2000) Proc. Natl. Acad. Sci., USA 97: 2235