Journal of Computer-Aided Molecular Design, 6 (1992) 113–130
ESCOM

J-CAMD 150

Computational simulations of the conformational behaviour of the adhesive proteins RGDS fragment

M. Cotrait[a,*], M. Kreissler[b], J. Hoflack[c], J.-M. Lehn[d] and B. Maigret[d]

[a]*Laboratoire de Cristallographie et Physique Cristalline, URA CNRS No. 144, Université de Bordeaux I, 351, Cours de la Libération, 33405 Talence Cedex, France*
[b]*Laboratoire de Physicochimie Théorique, URA CNRS No. 503, Chimie Physique A, 351, Cours de la Libération, 33405 Talence Cedex, France*
[c]*Marion-Merrell Dow Research Center, 16, Rue d'Ankara, BP 447 R/9, 67009 Strasbourg Cedex, France*
[d]*Laboratoire de Chimie Organique Physique, URA CNRS No. 422, Institut Lebel - Université L. Pasteur, 4, rue Blaise Pascal, 67000 Strasbourg, France*

Received 28 June 1991
Accepted 24 September 1991

Key words: Adhesive proteins; RGDS fragment; Adhesins; Integrins

SUMMARY

Many adhesive proteins present in extracellular matrices and in blood contain the tetrapeptide sequence -Arg-Gly-Asp-Ser- (or RGDS) at their cell recognition site. Since this sequence, or similar ones, was found in many proteins involved in major biological mechanisms, conformational investigations were performed on the RGDS fragment. A preliminary review of available crystal structures indicates that the RxDy sequences exhibit 3 well-defined structural patterns: one corresponding to a strong interaction between the Arg and Asp ionic side chains which are only about 4 Å apart, one with the ions separated by about 8 Å, and another in which the side chains are further apart (about 11 Å).

The conformational behaviour of the isolated RGDS fragment was next tackled using sequential building, Monte Carlo and molecular dynamics computational techniques. Analysis of the RGDS sequence conformational possibilities, as simulated in vacuum and in water solution, indicates that they can be classified into several conformational classes, which correspond roughly to the behaviour of the RGDS fragment as observed in protein matrices. This suggests the possibility of understanding the biological role of the RGDS or parent sequences in recognition processes.

INTRODUCTION

Since the discovery in the structure of fibronectin of the -Arg-Gly-Asp-Ser- (RGDS) sequence which has been proven to be responsible for the adhesive properties of this protein, similar sequences have been found in a surprisingly large number of extracellular proteins, and the versatility of the RGD-adhesion system is now widely accepted. The superfamily of proteins involved in

* To whom correspondence should be addressed.

0920-654X/$ 5.00

cell–cell and cell-extracellular matrix recognition has been called the adhesin family. Adhesins are associated with another superfamily of proteins, designated as integrins, which are mainly membrane glycoproteins. Some of these integrins have been isolated, identified and cloned (for reviews see [1–7]).

Apart from its usual action in the integrin family of cell surface receptors where the RGD sequence could be important in providing positional signals, the biological role of RGD-like sequences in several other biological systems of fundamental interest can be questioned. As examples: an Arg-Gly-Asp sequence within thrombin promotes endothelial cell adhesion [8]; the same sequence has been implicated in the mechanism of attachment of filamentous hemagglutinin to eukariotic cells [9]; an RFDS sequence is found in class I and II HLA antigens and might be involved in the cellular contact process occurring during the presentation of the antigen to the T lymphocytes [10]; this sequence is also supposed to play a role in host–parasite interactions [11] or in the rhinovirus action [12]; an RADS sequence is also found in the T-cell surface glycoprotein CD4, in the key region which is mainly responsible for the AIDS virus envelope gp_{120} protein recognition [13].

In most cases, the recognition process could be driven by a specific electrostatic interaction for which a particular disposition of the Arg...Asp ion pair is required, due to a particular conformation of the RGDS fragment inside its protein matrix and/or at the interface with its surrounding medium. Thus, it seems particularly relevant to investigate the conformational behaviour and properties of this peptidic fragment in the gas phase, in water solution, and in its protein environment. For this purpose we first analyzed the particular conformational characteristics of the RGDS-like sequences found in proteins of known crystal structures. We next undertook a conformational analysis of this sequence in vacuo by several computational methods. These results were finally compared to simulations where the solvent effects were explicitly taken into account.

METHODS OF CALCULATION

Taking account of the various advantages/failures of the current methodologies for conformational searching of complex molecules [14–16], we decided to investigate the conformational behaviour of the RGDS molecule by sequential building and Monte Carlo sampling followed by molecular dynamics simulations. The combination of all the preceding methods assured us of having a good representation of the conformational behaviour [17]. To mimic the pH range of the biological medium, the Arg and Asp residues were always considered in their charged side-chain states. An acetyl group and a methyl amide group were introduced at the N- and C- termini of the RGDS peptide to avoid undesirable electrostatic effects (Fig. 1). All dihedral angle (Φ,ψ,χ) values are given according to the IUPAC-IUB conventions [18]. The SYBYL software [19] was used for graphical display of the molecular systems.

Sequential building method

This method which combines low-energy conformations of the component parts of a peptidic molecule can be applied to a small molecule such as RGDS, since the number of all possible combinations to be considered in this stepwise building procedure is tractable in this particular case.

The intramolecular energies of all RGDS fragments were calculated using the ECEPP program [20], taking into account the modifications proposed for ECEPP83 [21]. The influence of water

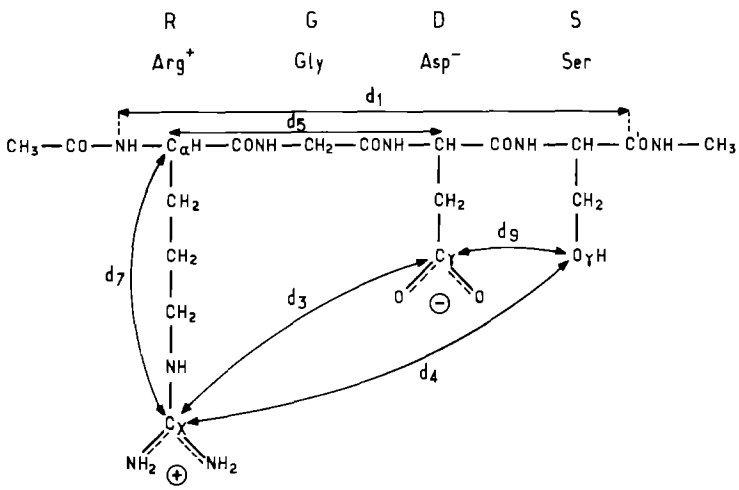

Fig. 1. The RGDS molecule model and the selected interatomic distances retained for clustering analysis.

was introduced here at a very crude level using variations of the dielectric constant as a function of interatomic distances. No explicit solvation energy terms were added to the usual potential energy function used during the building process. Such an effect [22] was introduced only during the final refinements of the whole RGDS templates.

The conformations selected for the individual residues were labelled according to the Liquori notation [23], and the corresponding low-energy states retained for the backbone of all individual Arg, Gly, Asp and Ser residues are presented in Fig. 2. The side-chain mobility was taken into account by mapping the side-chain conformational space with a 30° increment for all χ dihedral angles. As concerns the Ser residue, structures involving hydrogen bonding between the seryl $O^{\gamma}H$ hydroxyl and the seryl backbone as proposed by Lewis et al. [24] were discarded.

As a first step, stable conformations of the individual Arg and Gly residues were combined to give starting conformers of the Arg-Gly dipeptide. The Arg-Gly conformers built this way and presenting high conformational energies were eliminated, as high-energy conformations seldomly converge towards stable structures [25,26]. The energy of the remaining conformers was next minimized using the dihedral angles as the only variables. The same process was repeated to generate starting conformations for the RGD tripeptide. For this purpose, the stable conformers obtained previously for the Arg-Gly dipeptide were combined with those of the individual Asp residue and later refined. This procedure was repeated until low-energy conformations of the whole RGDS sequence were generated.

Metropolis sampling and cluster analysis

A larger sample of conformations can be generated by Monte Carlo techniques using the Metropolis algorithm [27]. This procedure has previously been used for conformational simulations [28,29] and has proved to be an efficient way of exploring energy hypersurfaces of medium complexity [30]. The conformational energy terms used here come from Scheraga et al. [31]. The procedure works as follows: a starting conformation **i** is chosen for the peptide by picking randomly

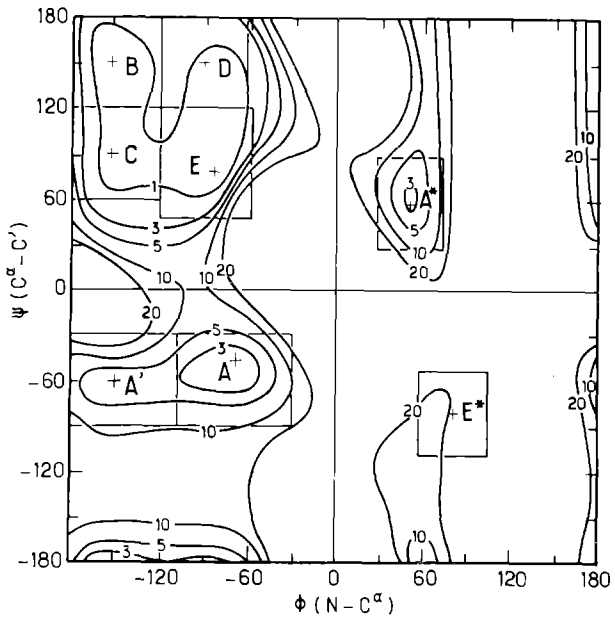

Fig. 2. Position on the usual Ramachandran plot of the individual amino acid conformations selected in the sequential sampling method. The conformational states selected for each individual residue are: more-or-less stretched structures (labelled B, C, D on the Ramachandran plot of Fig. 2) corresponding respectively to the following couples of ϕ, ψ dihedral angles: ($-150°$, 150°), ($-70°$, 150°), ($-150°$, 70°); the right- and left-handed α helices A ($-60°$, $-50°$) and A* (60°, 40°) respectively; the γ-turn E ($-80°$, 80°) and reverse γ-turn E* (80°, $-80°$).

selected (ϕ,ψ) dihedral angles in the Ramachandran energy map of each individual residue. The associated energy E_i is then calculated. This conformer is modified by selecting at random a new (ϕ,ψ) couple of dihedral angles in the Ramachandran map of a randomly selected residue. The energy E_j of the resulting conformer **j** is then calculated:

- if $E_j < E_i$, then conformation **j** is retained as the new conformational state and the process is repeated
- if $E_j > E_i$, the **j** conformer can nevertheless be retained according to its Boltzmann factor; otherwise it is rejected and the new conformational state is again **i**.

The process is a Markov chain which should converge toward high-probability conformational states as this procedure should prevent the molecule from being trapped in one of the numerous local energy minima of the conformational hypersurface. In the present calculations, 50 starting conformers were randomly selected, and each of the 50 corresponding Markov chains was improved during 2 000 steps, so that 100 000 conformers were thus considered to sample the RGDS energy space. During the sampling procedure, several interatomic distances of interest (see Fig. 1 for definition) were calculated and stored for each new conformational state selected. The classification of all the retained conformations into families of similar aspects was performed using these interatomic distances to obtain a similarity matrix between conformers. Two multidimensional statistical analysis methods of the conformational sample were used:

- clustering algorithms based on a hierarchy [32] in which the conformers are gathered in a step-

wise fashion to give several sets of increasing size until a unique set is reached; several possibilities exist to link conformers together until stable families are obtained: single, average, complete linkage methods, or centroid methods; all of them were used in the present work
- partitioning methods [33] in which a criterion based on the optimal separation into families is optimised in a stepwise fashion; the SAS statistical package [34] was used for this purpose.

Molecular dynamics simulations

The conformations generated and accepted from the previous sequential building and Monte Carlo sampling as the most representative of each conformational family obtained after clustering analysis were further shaken during a 100 ps molecular dynamics simulation at 300 K. The AMBER 3.0 package [35] was used during this stage. The bonds and bond angles are here treated as harmonic springs, and a torsional term is associated with the dihedral angles. The interactions between atoms separated by at least 3 bonds are described by a pairwise additive 1–6–12 potential, the parameters of which are those associated with the AMBER package in the full atom representation. In contrast to the sequential building and Metropolis methods which both used dihedral angles as variables, the conformational space is here explored using Cartesian coordinates. In this case, as the AMBER force field gives results similar to those obtained with ECEPP [36] and the Monte Carlo procedure, no large deviations in the minima due to different force fields were expected. The behaviour of the selected RGDS conformers was first studied in vacuo and next in a bath of water molecules.

To mimic the solvent effects, each conformer was placed at the center of a cubic water box of $30 \times 30 \times 30$ Å in which 1728 TIP3P water molecules [37] were initially considered. Any water with an oxygen distance less than 2.4 Å or with a hydrogen distance less than 2.0 Å was discarded. Such a removal of the solvent molecules at contact left about 700 water molecules, depending on the RGDS conformation. Periodic boundary conditions were used with a residue-based cutoff of 12 Å, a dielectric continuum ($\varepsilon = 1$), and a temperature of 300 K. No counterions were included in these simulations. Each conformer retained after the clustering procedure was first hydrated this way and the system peptide/solvent refined by molecular mechanics optimization. For this purpose, a steepest descent algorithm was used during 500 steps, after which conjugate gradients were switched on until convergency. Next, the systems were equilibrated for 10 ps of molecular dynamics starting with random velocities. This was followed by 100 ps of classical MD (no pressure monitoring) in which the total energy (kinetic + potential) is conserved. A time step of 0.0005 ps was used in the Verlet algorithm. The nonbonded interactions were processed by using a list-based method, and the lists were updated every 100 time steps. The SHAKE constraint procedure [38] was applied on all bond lengths. The coordinates of the entire peptide/solvent system were stored every 0.05 ps during each simulation. Each peptide/solvent simulation required about 100 h of CPU time on an IBM-3090-600 VF. For the simulations in vacuo, a similar procedure was used, the only difference being the use of a distance-dependent dielectric constant.

Although molecular dynamics simulations are able to sample conformational transitions in solution [39], in such 100 ps trajectories at 300 K, the friction forces due to the solvent may restrict the molecular movements to the conformational space around the starting conformers. A 500 ps simulation performed on one conformer [40] did not significantly change the situation described here, since we were interested in the mediation of electrostatic effects by water.

TABLE I
STRUCTURAL FEATURES FOR THE RxDy SEQUENCES FOUND IN THE BROOKHAVEN CRYSTALLOGRAPHIC PROTEIN DATA BANK

PDB reference name	Sequence	Arg...Asp C^{α}–C^{α} distance (Å)	Arg...Asp C^{β}–C^{β} distance (Å)	Arg...Asp C^{x}–C^{γ} distance (Å)
2ALP	Arg-Gly-Asp	5.575	7.151	11.403
1GCR	Arg-Gly-Asp	7.230	9.032	7.675
7TLN	Arg-Gly-Asp	6.206	8.344	14.013
2ATC	Arg-Asp-Asp	5.432	7.186	11.570
1FB4	Arg-Asn-Asp	6.179	6.109	4.334
1GP1	Arg-Asn-Asp	5.634	7.202	11.028
1IG2	Arg-Asn-Asp	6.097	6.324	4.781
1MBS	Arg-Asn-Asp	5.401	7.059	10.841
2CYP	Arg-Glu- Asp	5.878	7.760	12.852
1CYC	Arg-Gln-Asp	5.393	7.094	11.423
1MB5	Arg-Lys-Asp	5.344	7.049	11.588
1MBD	Arg-Lys-Asp	5.495	7.529	11.623
3MBN	Arg-Lys-Asp	5.642	7.560	12.279
1MBO	Arg-Lys-Asp	5.538	7.556	11.664
2MDH	Arg-Lys-Asp	5.277	7.118	11.867
4RLX	Arg-Lys-Asp	5.366	7.270	11.943
2PRK	Arg-Tyr-Asp	5.682	7.841	10.743
3GRS	Arg-His-Asp	6.940	9.156	13.184
1MEV	Arg-Phe-Asp	4.348	5.153	9.623
4RHV	Arg-Phe-Asp	7.343	8.923	8.357
4ATC	Arg-Leu-Asp	7.223	9.399	14.134
2TBV	Arg-Leu-Asp	6.393	7.330	10.344
1FX1	Arg-Ile-Asp	6.601	7.177	12.542
2LZM	Arg-Ile-Asp	5.291	7.332	11.086
1PHH	Arg-Ile-Asp	6.293	6.375	4.497
3ADK	Arg-Val-Asp	5.337	4.892	7.549
1CTX	Arg-Val-Asp	7.100	7.622	9.500
2DHB	Arg-Val-Asp	6.142	7.349	12.691
1FDH	Arg-Val-Asp	6.246	7.548	12.865
2HCO	Arg-Val-Asp	6.296	7.642	12.283
4HHB	Arg-Val-Asp	6.283	8.112	13.383
1HHO	Arg-Val-Asp	6.343	7.547	11.984
2MHB	Arg-Val-Asp	6.365	7.824	11.928
3PGK	Arg-Val-Asp	7.172	9.024	7.829
6PTI	Arg-Pro-Asp	6.568	8.158	10.601
1CC5	Arg-Ala-Asp	5.097	6.540	11.270
1CCR	Arg-Ala-Asp	5.525	7.455	11.490
1FBJ	Arg-Ala-Asp	6.741	8.102	8.857
2MCP	Arg-Ala-Asp	6.367	7.889	7.996
3WGA	Arg-Ala-Asp	6.448	8.889	12.990
2SNS	Arg-Thr-Asp	7.108	9.113	13.499
7ADH	Arg-Ser-Asp	4.889	6.273	10.237
1HMG	Arg-Ser-Asp	6.979	8.120	11.427
1MLP	Arg-Ser-Asp	5.318	7.043	9.094
1CD4	Arg-Ala-Asp	6.813	8.016	12.828
2HLA	Arg-Phe-Asp	6.468	5.704	4.607
2HLA	Arg-Val-Asp	5.327	7.205	9.432
2HLA	Arg-Thr-Asp	6.307	5.704	4.607

RESULTS

Experimental conformations analysed from protein data bank

While RGDS-like sequences (RxDy which differ from the original RGDS sequence by substitution of the Gly or Ser residues) can be widely found in sequence data banks (about 1200) and are present in proteins of fundamental interest, there are only about 50 crystal structures presenting such RxDy fragments which can be obtained from the Brookhaven protein data bank [41] (listed in Table 1).

When comparing all the RxDy fragments extracted from these protein matrices, the most important feature found concerns the relative positions of the Arg and Asp side chains (for example, from the calculation of distances between the C^{z} atom of the guanidinium side chain and the C^{γ} atom of the Asp carboxylate) which define 3 major classes of conformations for RxDy. In these classes the distance between the charged groups in Arg and Asp residues are respectively about 4, 7–8, and 11.5 Å (see the histogram of Fig. 3), corresponding respectively (i) to conformations with a salt bridge between Arg and Asp side chains, (ii) to conformations where the charged side chains are parallel and pointing towards the same direction or antiparallel with opposite directions, and (iii) to conformations with an elongated geometry between the guanidinium and carboxylate groups. In the first case, which is favored by strong electrostatic attractive interactions between the opposite guanidinium and carboxylate charged groups, both side chains may be found inside the protein globule or exposed at its surface. In the second one, the Arg and Asp side chains are both exposed to the protein outside. When the Arg and Asp side chains point in the same direction, one or two water molecules may be engaged between them and stabilized by a network of intermolecular hydrogen bonds. In the last case, one of the charged side chains points outside the protein while the other is found inside and is engaged in a network of several intramolecular H-bonds. To illustrate these situations, we have retained the following proteins:

(i) in the first case (salt bridge between the Arg...Asp side chains):

- immunoglobulin FAB (2FB4) with the RNDS sequence
- oxydoreductase (1PHH) with the RIDL sequence
- class I antigen (2HLA) with the RFDS and RTDA sequences

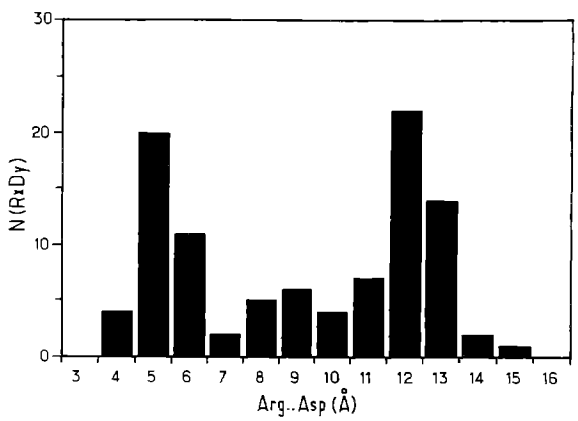

Fig. 3. Histogram showing the frequency of the Arg and Asp side chains in protein X-ray structures depending on distance intervals.

(ii) in the second case (Arg...Asp distances around 7–8 Å):
- gamma-II crystalline protein (1GCR) with the RGDY sequence
- rhinovirus (4RHV) with the RFDS sequence
- immunoglobulin (2MPC) with the RADS sequence

(iii) in the third case (Arg...Asp distances greater than 10 Å):
- thermolysin (7TLN) with the RGDG sequence
- alfa-lytic protease (2ALP) with the RGDS sequence
- influenza virus haemaglutinin (1HMG) with the RSDA sequence
- tomato bushy stunt virus (TBV) with the RLDL sequence
- cardio picornavirus coat protein (1MEV) with the RFDN sequence
- T-cell surface glycoprotein (1CD4) with the RADS sequence

In all these structures, the RxDy sequences appear near the protein surface and are located on β-strands or on loops connecting β-strands.

(1) The first group of RxDy sequences, which corresponds to the formation of a salt bridge between the Asp and Arg charged side chains, is found for example in the RND sequence of FB4 or in the RIG sequence of PHH protein (Fig. 4a).

(2) The second group of RxDy conformations is illustrated on one hand by the RGDY and RFDS sequences coming from the GCR and RHV proteins respectively. The conformations of these 2 fragments appear very similar (Fig. 4b). The Arg...Asp side-chain distance is about 8 Å and the 2 backbones have a similar folding with a **U** shape formed by the Arg side chain, the Gly backbone and the Asp side chain. The Arg and Asp side chains are almost parallel and point outside the surface of the protein, thus making possible strong electrostatic intermolecular interactions with the protein surroundings.

Another subgroup which illustrates the antiparallel relative position between the Arg and Asp side chains and the **S** shape of the side chain–backbone–side chain guanidinium-carboxylate graph is illustrated by the MPC protein fragment (Fig. 4c).

(3) In the third group composed of the remaining fragments, the Arg and Asp side chains are located further apart from one another (the distance being around 11 Å).

Two subgroups can be defined according to the backbone conformations and to the position of the Asp residue: in the first subgroup (see Fig. 4d) which contains the RGDG (TLN), RGDS (ALP) and RSDA (HMG) sequences, the Asp side chain is engaged inside a protein pocket making several hydrogen bonds with the surrounding residues or water molecules, while the guanidinium moiety points outside the surface; in the second subgroup coming from the TBV, MEV and CD4 proteins (see Fig. 4e), the Asp residue is accessible and seems free to interact with species outside the protein while the guanidinium is involved in intramolecular interactions.

Sequential conformational building

After the preliminary calculations performed with a dielectric constant of 2, the Arg^+ guanidinium and the Asp^- carboxylate groups were found to form a salt bridge in most of the stable conformers obtained, a situation already observed in NMR experiments on GRGDSG in DMSO [42–44]. But since such a situation is not the only one expected when the RGDS sequence is involved in protein–protein interactions, further studies were carried out with a higher value of the dielectric constant ($\varepsilon = 8$) as an attempt to simulate the hydrophilic biological medium [45]. In this case, 504 possible starting conformers were obtained for the RGD moiety following the combina-

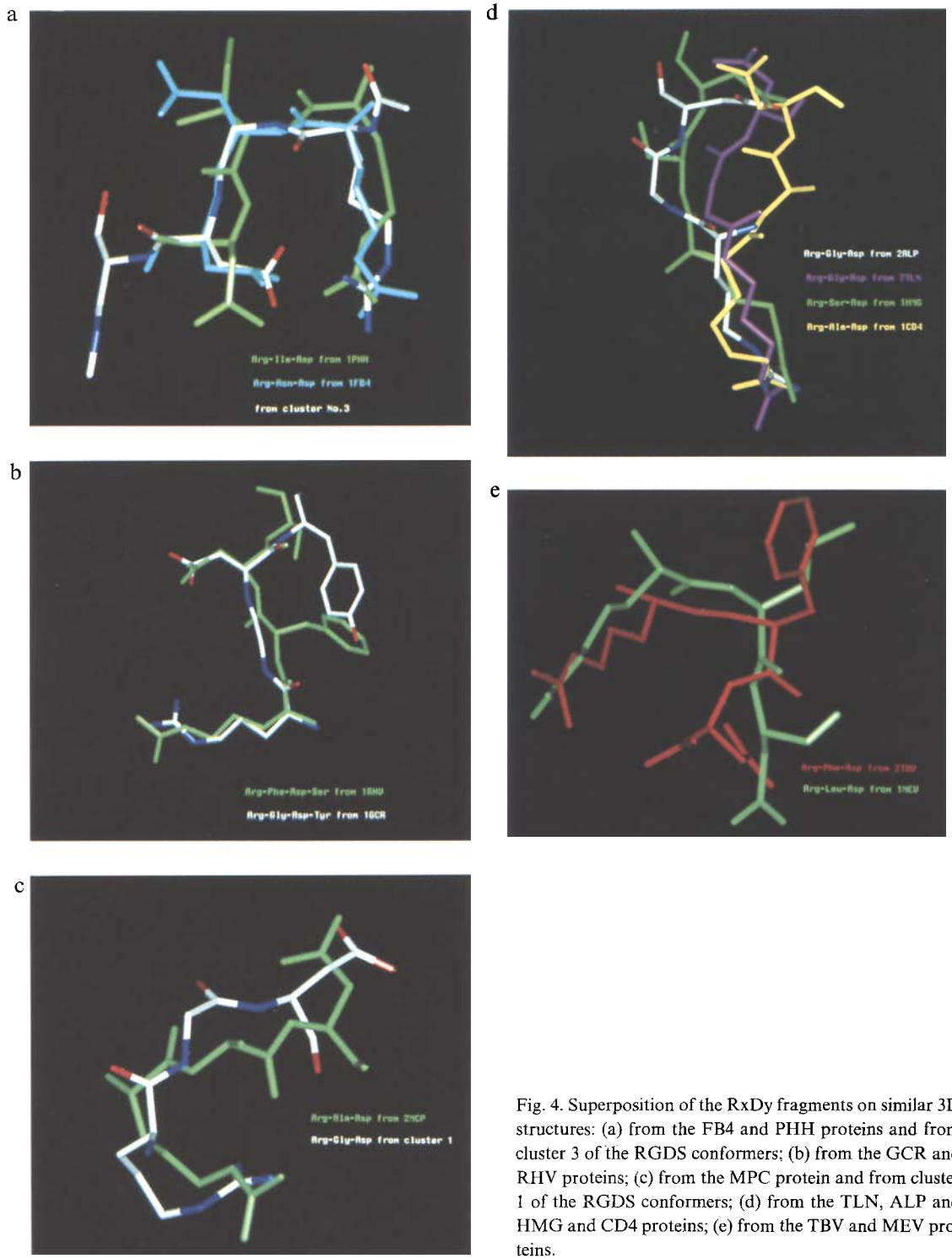

Fig. 4. Superposition of the RxDy fragments on similar 3D structures: (a) from the FB4 and PHH proteins and from cluster 3 of the RGDS conformers; (b) from the GCR and RHV proteins; (c) from the MPC protein and from cluster 1 of the RGDS conformers; (d) from the TLN, ALP and HMG and CD4 proteins; (e) from the TBV and MEV proteins.

tion sampling procedure. After energy refinement, only 117 conformations were retained within a 5 kcal/mol interval above the lowest-energy structure. They covered a broad range of conformational possibilities, but several of them presenting a γ or reverse γ turn around the Gly residue were found to be particularly stable. Combining these 117 most stable conformers of the RGD part to the most stable conformers of the individual Ser residue gave the starting sets of the RGDS peptide. After energy refinement, 63 conformations were found in the 5 kcal/mol interval above the lowest energy one. The 21 lowest-energy conformations (3 kcal/mol range) are presented in Table 2. In most of these conformers the Arg and Asp charged side chains were rather far apart ($C^x....C^\gamma$ distance > 6.5 Å, with a mean value of 7.5 Å). Nevertheless, the most stable conformers always involved close favorable contacts between the guanidinium and carboxyl groups ($C^x....C^\gamma$ distance between 4–6 Å). In several of the conformers obtained, a turn around the glycyl residue corresponding to γ or γ-reverse turn structures was observed. These results reflect the large conformational flexibility of the RGDS molecule, as no particular conformation was significantly more stable than the others.

Metropolis conformational sampling and analysis

From the 100 000 conformations (50 × 2 000) generated by the Metropolis procedure, a set of 1981 *different* stable RGDS conformers was obtained. Each of these conformations was described

TABLE 2
CONFORMATIONAL DISTRIBUTION FOR RGDS PEPTIDE ACCORDING TO THE SEQUENTIAL BUILDING PROCEDURE (THE D_1 AND D_3 DISTANCES ARE DEFINED IN FIG. 1)

Conformation no.	Liquori conformational type	Arg ϕ,ψ values (°)	Gly ϕ,ψ values (°)	Asp ϕ,ψ values (°)	Ser ϕ,ψ values (°)	ΔE (kcal/mol)	d_3 distance (Å)	d_1 distance (Å)
1	AE*DA	−71, −50	88, −67	−65, 139	−70, 51	0.0	5.12	8.81
2	BAEA	−156, 121	−74, −77	−76, 108	−73, −39	0.2	6.29	9.06
3	BAAA	−160, 134	−71, −44	−52, −49	−65, −52	0.4	4.05	7.48
4	CAEE	−157, 117	−71, −72	−79, 94	−85, 81	0.8	6.23	9.60
5	BAAE	−158, 137	−70, −50	−55, −51	−102, 72	1.0	4.25	4.43
6	CADB	−159, 113	−79, −73	−72, 123	−153, 148	1.1	6.67	10.05
7	DEAA	−72, 123	−88, 73	−69, −44	−71, −43	1.4	4.50	10.92
8	CAEA	−158, 111	−72, −66	−80, 91	−71, −45	1.5	7.64	8.92
9	AE*EE	−71, −49	83, −83	−81, 102	−84, 82	1.7	6.72	10.03
10	AE*BA	−73, −43	104, −58	−154, 139	−74, −47	1.8	7.39	9.47
11	AAAA	−69, −42	−71, −37	−65, −43	−66, −47	1.8	12.32	7.58
12	BEAA	−157, 125	−88, 73	−70, −43	−70, −40	2.0	4.54	13.17
13	DEDA	−72, 123	−89, 74	−71, 142	−69, −41	2.1	4.57	11.63
14	CA*DA	−157, 118	−79, −111	−68, 134	−72, −42	2.3	8.15	9.71
15	AAAB	−68, −39	−66, −40	−67, −44	−156, 124	2.4	12.38	4.11
16	A*AEA	57, 92	−63, −56	−82, 97	−70, −40	2.6	6.55	8.86
17	AE*BE	−72, −43	103, −59	−153, 141	−82, 86	2.7	7.28	10.42
18	AA*AB	−71, −48	70, −120	−77, −43	−159, 157	2.8	8.67	5.87
19	AAAE	−70, −41	−65, −43	−65, −46	−84, 79	2.9	13.11	5.58
20	DAAD	−71, 132	−161, −60	−63. −47	−72, 121	2.9	9.97	5.72
21	AE*EB	−71, −47	88, −72	−73, 108	−159, 148	3.0	5.79	9.59

by the set of 10 interatomic distances defined in Fig. 1, so that the whole conformational space was here described by a (1981 × 10) matrix. Thus each conformation was a point in a 10-dimensional space.

The first step consists here in visualizing the shape of the cloud formed by the 1981 points in this complicated space, and in reducing the dimensions of this space if possible. The Principal Component Analysis (PCA) method and elementary statistics are suitable for this purpose [32]. Such an analysis shows that the structural information contained in the cloud is essentially described (90%) by the first 6 principal axes. This means that some of our 10 variables are strongly correlated and that the cloud presents well-separated groups. For example, when looking at the mean values of the N- to C-terminal distance and of the $C^{x}(Arg^{+})...C^{\gamma}(Asp^{-})$ distance respectively, it is possible to distinguish the extended and folded forms of the peptidic backbone, as well as the relatively close and remote positions between the Arg and Asp side chains.

The determination of the optimal number of representative clusters is not a simple problem as no unique answer exists concerning the best partition of the cloud into groups. It is thus advisable to compare different classification methods and criteria to obtain a realistic evaluation of this number of classes. Following the procedure of Metropolis et al. [27], our cloud of conformers separated well into 3 clusters of structures containing respectively:

- 620 conformers (31% of the sample) for cluster 1
- 793 conformers (40%) for cluster 2
- 568 conformers (29%) for cluster 3

Moreover, cluster 2 consists of subgroups a and b representing respectively 380 and 413 conformers. This distribution does not correspond to a thermodynamic equilibrium, but to the configurations accepted during the simulation and retained for analysis. Table 3 shows that, depending on the cluster, extended and folded backbone conformations were identified roughly in the same proportions. These results agree quite well with those obtained by the sequential sampling procedure. The SNOOPI software was used to represent the stereoviews of the 4 representative conformers [46]. It appears that the more compact conformations are found in cluster 1 while the more extended are in cluster 2a; the extended backbone conformations presenting close proximity between the Arg and Asp side chains are found in cluster 3, while folded backbone structures with remote Arg and Asp side chains are in cluster 2b. The most representative conformations of each cluster (obtained by choosing and then refining, in each cluster, the conformation closest to the center of the cluster) are presented in Fig. 5a, and by the Ramachandran map of Fig. 5b. Such a map shows that the Arg and Ser residues generally adopt stretched structures while the Gly and

TABLE 3
AVERAGES OF THE END-TO-END, AND GUANIDINIUM...ACETATE DISTANCES FROM THE METROPOLIS SAMPLING

Weight in the sample (%)	Cluster no.	N-C terminal distance (Å)	$C^{x}_{(Arg)}...C^{\gamma}_{(Asp)}$ distance (Å)
31	1	6.8	8.3
19	2a	9.3	9.5
21	2b	4.3	10.8
29	3	10.1	5.1

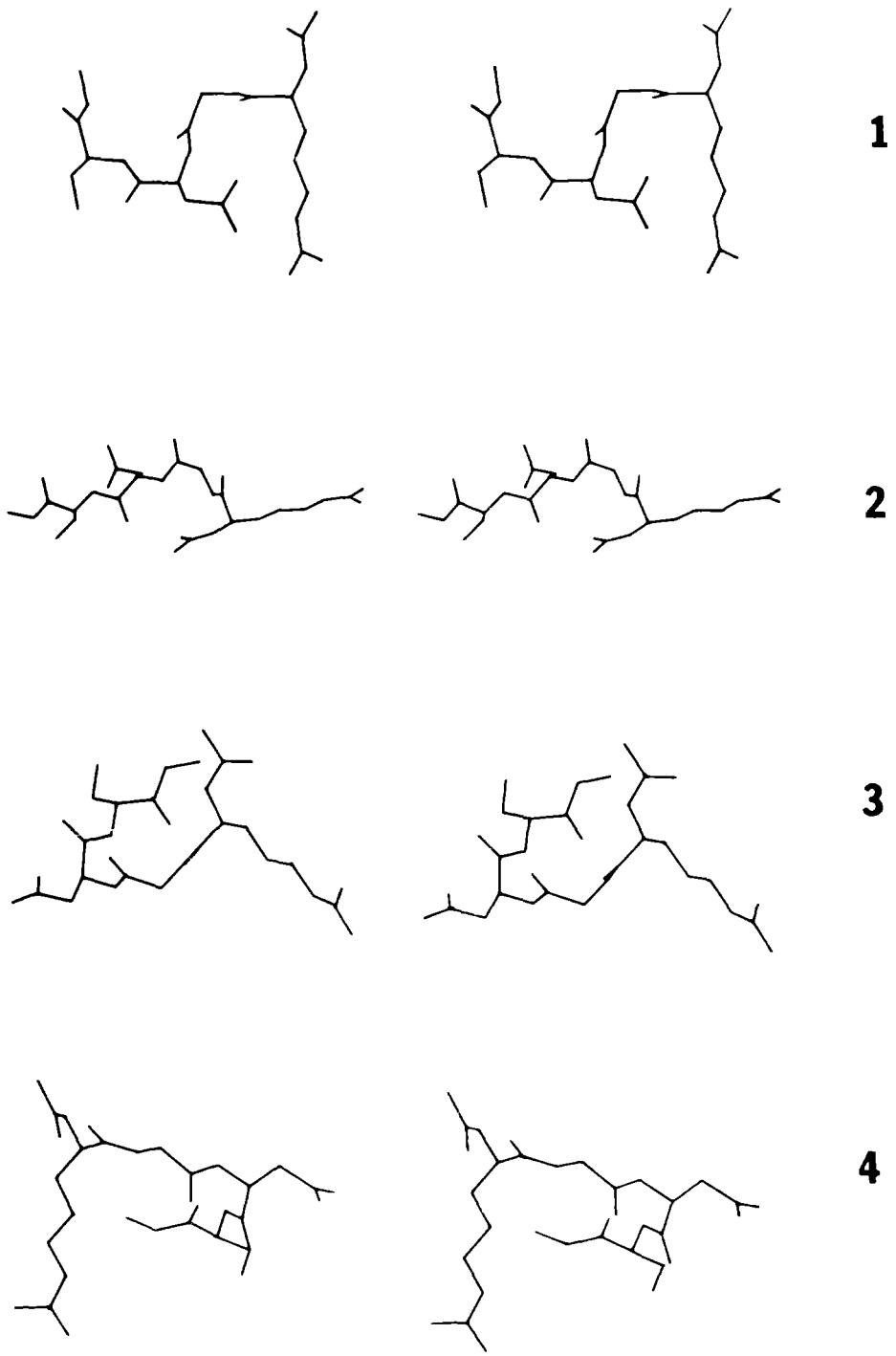

Fig. 5. (a) Stereo view of the four representative model conformers obtained from the clustering analysis.

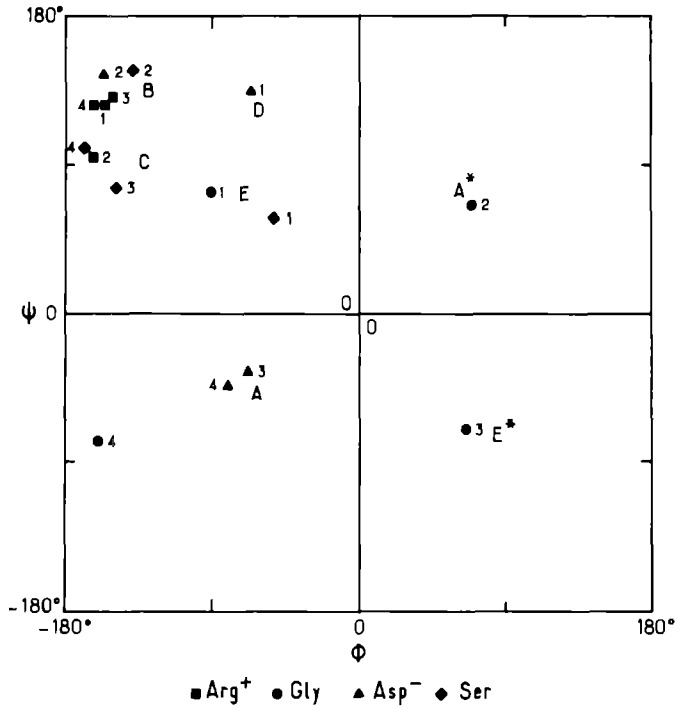

Fig. 5. (b) Corresponding (ϕ,ψ) points of the four model conformers (shown in Fig. 5a) on the usual Ramachandran map.

Asp residues present more possibilities, in some agreement with the backbone structures found in the RxDy fragments of the X-ray protein structures examined above. These protein fragments can be connected to the classification proposed from this clustering analysis: the first group of X-ray RxDy structures is related to cluster 3, the second group to cluster 1, and the third to clusters 2a,b. Such a result suggests that the conformations observed for the RxDy fragments in proteins may also be related to stable structures of the isolated RxDy molecule alone, which are not deeply modified by the surrounding residues in the macromolecules. This may suggest that the RxDy sequence acts as a template which has been retained by evolution because of its original intrinsic conformational behaviour. There is a rather good correlation between sequential and Monte Carlo results, as shown in Tables 2 and 3.

Many conformers with average distances of 6.5 Å and 9.5 Å for d_1 and d_3 respectively are in agreement with cluster 1. Only 3 conformers can be related to cluster 2a, with mean distances d_1 and d_3 respectively close to 7.9 and 9.4 Å. A few conformers have a long d_3 distance associated with a short d_1 distance and may correspond to cluster 2b.

Finally, conformers with a short d_3 distance constitute about one third of the sequential distribution, with an average d_1 distance close to 10.3 Å. They can be related to cluster 3.

Molecular dynamics calculations

Molecular dynamics simulations were performed in order to investigate the conformational stability of the representative conformers of the 4 families obtained above, and to look for possible

TABLE 4
GEOMETRICAL CHARACTERISTICS OF THE $C^{x}_{(Arg^{+})}...C^{y}_{(Asp^{-})}$ DISTANCE AFTER THE 100 ps DYNAMICS SIMULATIONS FOR EACH OF THE RETAINED REPRESENTATIVE CONFORMERS

Conformation from cluster no.	$C^{x}...C^{y}$ distances (Å)			
	Mean value	Minimum	Maximum	Starting value
1 in vacuo	6.01	3.36	8.01	6.96
2 in vacuo	3.96	3.31	11.91	11.64
3 in vacuo	3.73	3.43	4.02	3.62
1 in water	7.07	6.36	8.17	7.58
2 in water	7.32	5.23	11.45	10.50
3 in water	4.02	3.30	4.45	3.63

interconversions between them. Thus, each of these conformers was observed during a 100 ps trajectory preceded by 10 ps of equilibration of the system at 300 K.

The results in vacuo are, as expected, driven mainly by the electrostatic attraction between the charged Arg and Asp side chains, so that all 3 conformations for which these side chains are far apart (clusters 1 and 2a,b) present a conformational transition, making them return more-or-less rapidly to cluster 3 characterized by a salt bridge between the charged Arg and Asp groups. This situation is clearly depicted on the plots of Figs. 6a–c presenting the evolution of the $C^{x}(Arg^{+})...C^{y}(Asp^{-})$ distance during the 100 ps trajectories for the conformers of clusters 1, 2b and 3, respectively. It appears in Fig. 6c that the initial salt bridge existing in cluster 3 between the charged guanidinium and carboxyl groups was conserved during the simulation. The conformation of cluster 1 (which has been related to one of the most interesting RxDy protein fragments), with a starting Arg...Asp side-chain distance around 8 Å, gave a transition after 70 ps to a conformation close to that observed in cluster 3. The same occurred even faster (after 10 ps) for the conformation of cluster 2b. This means that the extended relative positions of the Arg and Asp side chains were not stable enough to characterize the conformational behaviour of the RGDS system in vacuo or in nonpolar solvents. In this case, the strength of the ionic interaction between the charged groups was large enough to drive the equilibrium towards cluster 3 conformations, where the salt bridge between the Arg and Asp ion pair was dominant.

However, such a behaviour might differ in water, as hydration of the charged groups may decrease the effect of the electrostatic forces. The dynamics of the same 4 conformers was therefore simulated in an explicit bath of water molecules, and the curves depicting the evolution of the distance between the charged Arg and Asp groups plotted again.

The fluctuation of the Arg...Asp distance is illustrated in Fig. 7b for the conformer of cluster 2b which was rapidly transformed in vacuo into a conformation of cluster 3. Then the starting distance of 11 Å decreased rapidly to appear stabilized around 7 Å. In such a situation, the molecule appeared to be stabilized in a conformation very similar to that of cluster 1. When the molecule was hydrated, the organization of the water was strong enough to balance the electrostatic attraction between the charged groups and then to maintain the molecule in stable structures where the salt bridge was not the driving force. This result highlights the importance of cluster 1 conformers

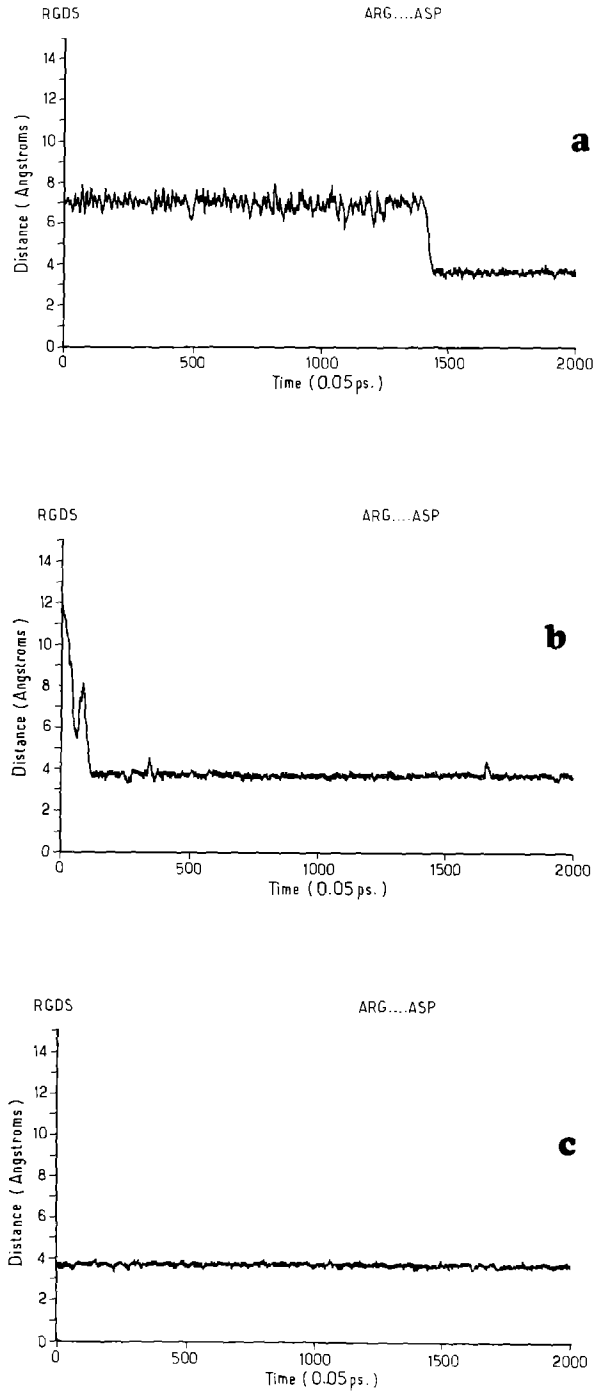

Fig. 6. Evolution of the $C^{\alpha}_{(Arg^{+})}...C^{\gamma}_{(Asp^{-})}$ distance during molecular dynamics simulations in vacuo. (a) Starting conformer is taken from cluster 1. (b) Starting conformer is taken from cluster 2b. (c) Starting conformer is taken from cluster 3.

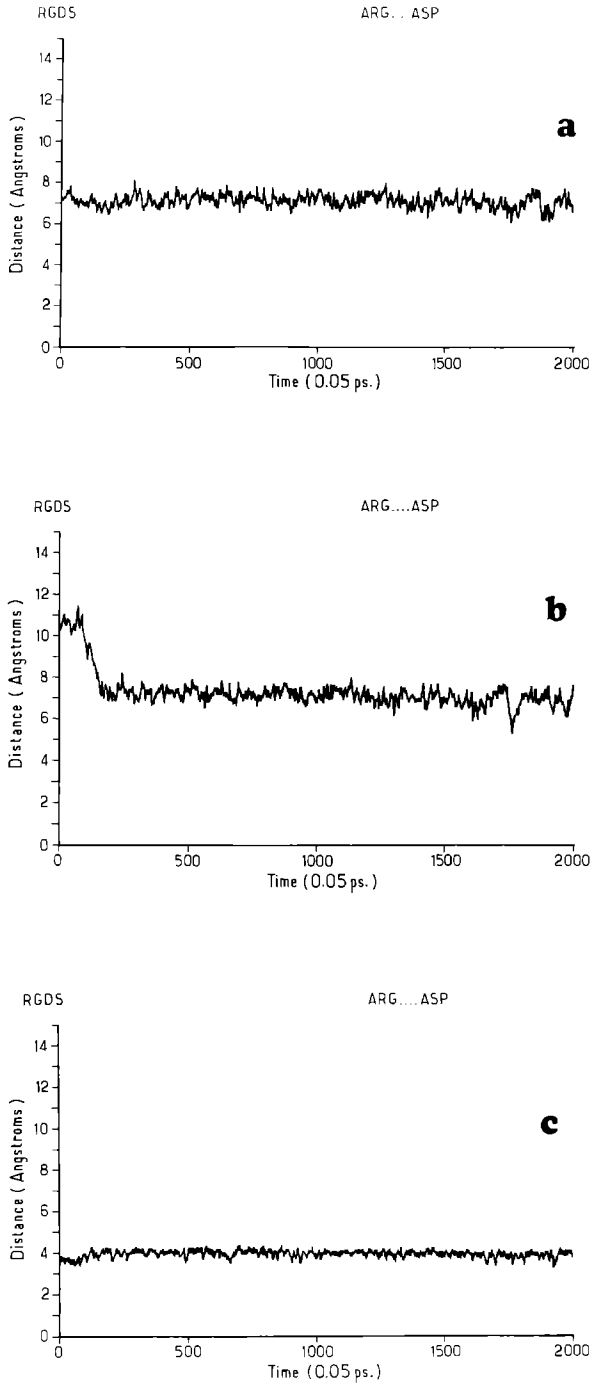

Fig. 7. Evolution of the $C^{x}_{(Arg^{+})}...C^{y}_{(Asp^{-})}$ distance during molecular dynamics simulations in water. (a) Starting conformer is taken from cluster 1. (b) Starting conformer is taken from cluster 2b. (c) Starting conformer is taken from cluster 3.

which appear, from these calculations on the RGDS molecule in water, to be the only stable conformational possibility besides the salt bridge ones.

CONCLUSIONS

The protein databank survey has shown that RxDy sequences in protein crystals mainly adopt 3 kinds of structures which appear as the most stable possibilities of the RGDS molecule itself. Environmental effects are of crucial importance for the conformational possibilities of the short sequence. Very recently, Eggleston and Feldman [47] published a crystal structure of the consensus sequence of RGD. Though the oligopeptide environment was different in their study since they crystallized the RGD molecule surrounded by 4 water molecules, when we performed our computer simulations with an isolated RGDS sequence or placed in a water bath, there exists a striking similarity between the experimental conformer and one conformer of cluster 1 which we have determined. In this case, we would like to stress two similar features:

(i) the bending of the molecule through the glycine residue ($\phi_2^{exp} = -85.5°$, $\phi_2^{cal} = -84°$)

(ii) the Arg and Asp side chains are on opposite sides of the peptide backbone ($d_{Arg...Asp}^{exp} = 9$ Å, $< d_{Arg...Asp} >^{cal} = 8.3$ Å) with an **S**-type configuration similar to that described in Fig. 4c.

The importance of the cluster 1 conformational type has already been asserted from molecular dynamics calculations in water. A conformational equilibrium between salt bridge conformers and those from cluster 1 should exist in solution or in proteins where the Arg and Asp side chains are not involved in strong intermolecular interactions. Such an equilibrium, monitored by water molecules which could replace or displace the guanidinium...carboxylate interactions, could be of crucial importance in several recognition processes involving the RGD sequence. Such cooperativity of the surface salt bridges has already been recognized to be of importance in molecular mechanisms of allosteric regulation [48]. Therefore, one could suggest that in addition to providing a site for strong electrostatic intermolecular interaction, the 'open' and 'closed' forms of the RxDy unit could also act in cooperation as an allosteric switch leading to larger changes in the overall shape of the macromolecule, so as to allow strong and specific interaction with remote parts of its partner over an extended molecular area. Blocking or inducing such a change by means of a specially designed binding molecule might thus repress or express the corresponding biological function. It thus seems possible to mimic the behaviour of this sequence involved in important biological events and to use the present results to design either synthetic analogues showing similar biological properties, or artificial receptor molecules that might bind to the RGDS fragment and stabilize one or other of the preferred conformations. Such work is in progress in our laboratories, as well as a detailed analysis of the RGDS/water interactions.

ACKNOWLEDGEMENTS

This work was supported by the Groupement Scientifique between IBM-France and the French National Research Center (CNRS).

REFERENCES

1 Pierschbacher, M.D., Ruoslahti, E., Sundelin, J., Lind, P. and Peterson, P.A., J. Biol. Chem., 257 (1982) 9593.
2 Hynes, R.O., Sci. Amer., 254 (1986) 32.

3 Ruoslahti, E. and Pierschbacher, M.D., Cell, 44 (1986) 517.
4 Ruoslahti, E. and Pierschbacher, M.D., Science, 238 (1987) 491 (and references therein).
5 Hynes, R.O., Cell, 48 (1987) 549.
6 Albelda, S.M. and Buck, C.A., FASEB J., 4 (1990) 2868.
7 Ruoslahti, E., J. Clin. Inv., 87 (1991) 1.
8 Bar-Shavit, R., Sabbah, V., Lampugnani, M.G., Marchisio, P.C., Fenton II, J.W., Vlodavsky, I. and Dejana, E., J. Cell Biol., 112 (1991) 335.
9 Relman, D.A., Domenighini, M., Tuomanen, E., Rappuoli, R. and Falkow, S., Proc. Natl. Acad. Sci. USA, 86 (1989) 2637.
10 Sherif, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finkel, B.C. and Davies, D.R., Proc. Natl. Acad. Sci. USA, 84 (1987) 8075.
11 Ouaissi, M.A., Cornette, J. and Capron, A., Mol. Biochem. Parasitol., 19 (1986) 201.
12 Auffray, C. and Novotny, J., Human Immunol., 15 (1986) 381.
13 Auffray, C., Piatier-Tonneau, D. and Kroemer, G., Trends Biotechnol., 9 (1991) 124.
14 Nayeem, A., Vila, J. and Scheraga, H.A., J. Comput. Chem., 12 (1989) 594.
15 Shin, J.K. and Jhon, M.S., Biopolymers, 31 (1991) 177.
16 Bruccoleri, R. and Karplus, M., Biopolymers, 29 (1990) 1847.
17 Howard, A.E. and Kollman, P.A., J. Med. Chem., 31 (1988) 1669.
18 I.U.P.A.C.–I.U.B. Commission on Biochemical Nomenclature, J. Mol. Biol., 52 (1970) 1.
19 TRIPOS Associates Inc., St. Louis, MO 63144, U.S.A.
20 E.C.E.P.P. Empirical Conformation Energy Program for Peptides (1976) Q.C.P.E. program No. 286, Bloomington, U.S.A.
21 Chuman, H., Momany, F.A. and Schafer, L., Int. J. Pept. Prot. Res., 24 (1984) 233.
22 Richards, R., J. Mol. Biol., 82 (1974) 1.
23 Liquori, A.M., Quart. Rev. Biophys., 2 (1969) 65.
24 Lewis, P.N., Momany, F.A. and Scheraga, H.A., Isr. J. Chem., 11 (1973) 121.
25 Anderson, J.S. and Scheraga, H.A., Macromolecules, 11 (1978) 805.
26 Fitzwater, S., Hodes, Z. and Scheraga, H.A., Macromolecules, 11 (1978) 805.
27 Metropolis, N.A., Rosenbluth, A.W., Teller, A.M. and Teller, E.J., J. Chem. Phys., 21 (1953) 1087.
28 Marchionini, C., Maigret, B. and Premilat, S., Biochem. Biophys. Res. Commun., 112 (1983) 339.
29 Kreissler, M., Pesquer, M., Maigret, B., Fournie-Zaluski, M.C. and Roques, B.P., J. Comput.-Aided Mol. Design, 3 (1989) 85.
30 Wille, L.T., Nature, 324 (1986) 46.
31 Scheraga, H.A., Adv. Phys. Org. Chem., 6 (1968) 103.
32 Massart, D.L. and Kaufman, L., The Interpretation of Analytical Chemistry Data by the use of Cluster Analysis, Chemical Analysis Monographs, John Wiley & Sons, New York, 1983.
33 Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, John Wiley & Sons, New York, 1973.
34 S.A.S. Institute Inc., Box 8000, CARY, NC 27511, U.S.A.
35 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
36 Roterman, I.K., Lambert, M.H., Gibson, K.D. and Scheraga, H.A., J. Biomol. Struct. Dyn., 7 (1989) 421.
37 Jorgensen, W., Chandresekhar, J., Madura, J., Impey, R. and Klein, M., J. Chem. Phys., 79 (1983) 926.
38 Ryckaert, J.P., Ciccotti, C. and Berendsen, H.J.C., J. Comput. Phys., 23 (1977) 327.
39 Tobias, D.J., Mertz, J.E. and Brooks III, C.L., Biochemistry, 30 (1991) 6054.
40 Hoflack, J. and Maigret, B., work in progress.
41 Bernstein, F.C., Koetzle, T.G., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rogers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 122 (1977) 535.
42 Genest, M., Marion, D., Caille, A. and Ptak, M., (1988) Second Forum on Peptides, Nancy (France).
43 Mayer, R. and Lancelot, G., J. Am. Chem. Soc., 103 (1981) 4738.
44 Negre, E., Marion, D., Roche, A.C., Monsigny, M. and Mayer, R., (1988) Second Forum on Peptides, Nancy, France.
45 Cotrait, M., Int. J. Pept. Prot. Res., 22 (1983) 110.
46 Davies, K., (1983) SNOOPI, Program for drawing crystal and molecular diagrams. Chemical Crystallography Laboratory, University of Oxford, U.K.
47 Eggleston, D.S. and Feldman, S.H., Int. J. Pept. Prot. Res., 36 (1990) 161.
48 Horovitz, A., Serrano, L., Avron, B., Bycroft, M. and Fersht, A.R., J. Mol. Biol., 216 (1990) 1031.