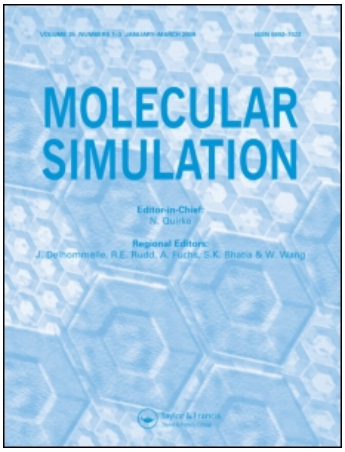

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Multivariate Investigations of Random-Coil and Ordered-Structure Conformations of the TYR181-to-TYR188 Segment of HIV-1 Reverse Transcriptase

Peter P. Mager[a]
[a] Research Group of Pharmacochemistry, Institute of Pharmacology and Toxicology of the University, Saxony, Germany

To cite this Article Mager, Peter P.(1997) 'Multivariate Investigations of Random-Coil and Ordered-Structure Conformations of the TYR181-to-TYR188 Segment of HIV-1 Reverse Transcriptase', Molecular Simulation, 19: 1, 17 — 41
To link to this Article: DOI: 10.1080/08927029708024136
URL: http://dx.doi.org/10.1080/08927029708024136

Molecular Simulation, 1997, Vol. 19, pp. 17–41
Reprints available directly from the publisher
Photocopying permitted by license only

MULTIVARIATE INVESTIGATIONS OF RANDOM-COIL AND ORDERED-STRUCTURE CONFORMATIONS OF THE TYR181-TO-TYR188 SEGMENT OF HIV-1 REVERSE TRANSCRIPTASE

PETER P. MAGER*

Research Group of Pharmacochemistry, Institute of Pharmacology and Toxicology of the University, D-04107 Leipzig, Härtelstr. 16-18, Saxony, Germany

(Received October 1996; accepted October 1996)

The segment Tyr181 to Tyr188 was dissected from the HIV-1 reverse transcriptase (RT). The segment contains two amino acids (Asp185, Asp186) of the catalytic aspartyl triad (Asp110, Asp185, Asp186) and two amino acids (Tyr181, Tyr188) of the nonnucleoside RT inhibitor (NNRTI) binding sites. Hydrogen-bonding forces between the folded peptide chain play the greatest role in specifying the folding pattern, an ordered Ω-loop. Minimum-energy conformations of a random coil of the segment, and of ordered structures (Ω-loop, 3_{10}-helix, left-hand and right-hand α-helix, antiparallel and parallel β-sheet, extended β-sheet, double-C7 γ-turn), were investigated using canonical correlation of subsets of Cartesian coordinates, and cluster and factor analysis. There is a hierarchy of geometrically similar conformers. The α-helix and Ω-loop form the first cluster. The π-helix and extended γ-turn, the 3_{10}-helix and left-hand α-helix, and the antiparallel and parallel β-sheets form the second cluster. The random coil forms an third cluster and is completely discriminated from the remaining subset of conformations.

Keywords: Molecular simulation; conformational similarity; HIV-1 reverse transcriptase; Ω-loop; cluster analysis; factor analysis

1. INTRODUCTION

To treat chemically the acquired immunodeficiency syndrome (AIDS) induced by the immunodeficincy virus type 1 (HIV-1), the viral reverse

*ppm@pharma.photo.uni-leipzig.de

transcriptase (RT) plays a key role. Active RT is a strikingly asymmetric heterodimer composed of two differently folded 66 kDa (p66) and 51 kDa (p51) subunits. The p66 subunit is a bifunctional enzyme containing a DNA polymerase and ribonuclease H (RNase H) activity. The polymerase and RNase activities cooperate to convert the single-stranded genomic RNA of HIV-1 into a linear double-stranded DNA that is subsequently integrated into host cell chromosomes[1–5]. During a single cycle of viral replication, HIV-1 RT catalyzes the incorporation of approximately 20,000 nucleotides. As RT does not appear to be essential in the normal functioning of mammalians, it is an attractive target for structure-based drug design. Present-day chemotherapeutic agents interfere with the polymerase activity of the enzyme.

Molecular modelling of the allosteric area of RT polymerase has shown two subregions (amino acid positions 98–106 and 179–190). In particular, the segment Tyr181 to Tyr188 is of interest[6–9], it forms a so-called Ω-loop[7–9]. The treatment of solvents led to little changes in conformation of the Ω-loop of RT because the overall shape was maintained. This was also valid for a water-solvated and desolvated state[8]. In contrast, simulation at body temperature and at energetically activated states yielded a loss of the ordered structure[9]. This focused interest on investigating the random-coil/ordered-conformation conversion. As the naturally occuring tertiary structure of proteins is formed along a relatively unknown, "largely mysterious" folding pathway, it should be useful to apply molecular simulation technique in order to reveal conformation changes.

2. MATERIAL AND METHODS

All-atom molecular mechanics (MM) methods[10–12] were used. The MM + empirical potential (force field) is an improved MM2/MM3 version[10, 11]. It was chosen as first method but it does not have an option for changing the dielectric constant ε, in contrast to AMBER force field[12] (Assisted Model Building and Energy Refinement). The whose MM + procedure was repeated with electrostatic parameters of a "quantum chemistry (QC) module" (an approximation of semiempirical QC applicable for medium-sized and large peptides), the connectivity-based iterative partial equalization of orbital electronegativity (Gasteiger method[13]). Then, the AMBER force field was employed.

Correlation-gradient geometry optimization[14] was achieved by the following steps: The structures were refined using a conjugate gradient minimizer

(Fletcher-Reeves modification of the Polak-Ribière method). Convergence was obtained when the gradient root mean square *RMS* was $RMS < 0.05$. The conformations were initially energy minimized using the MM+ force field without an electrostatic term. After including the partial charges, the resulting conformation contains the molecular electrostatic potential and electrostatic energies.

3. RESULTS

Data Source

The crystal structures of complexed and unliganded HIV-1 (wild-type) reverse transcriptase (E.C.2.7.7.49) were uploaded from the Brookhaven Protein Data Bank[16] by using Cartesian coordinates. The resolution range of the nevirapine-liganded RT (3HVT in PDB nomenclature) was 8.0–2.9 Å, of the unliganded RT (1RTI) it was 25.0–3.0 Å; the resolutions were sufficient to permit assignment of amino acid positions through-out the structure and allows subsequent molecular modelling. Note that nevirapine belongs to the class of nonnucleoside reverse transcriptase inhibitors (NNRTIs) that act noncompetitively. Measurement occured at 113 Kelvin temperature.

Choice of the Segment Tyr181 to Tyr188

Of particular interest are the amino acids in positions 181 to 188. The segment contains two amino acids (Asp185, Asp186) of the evolutionarily conserved catalytic aspartyl triad (Asp110, Asp185, Aso186) and two amino acids (Tyr181, Tyr188) of the allosteric resp. nonnucleoside RT inhibitor binding sites. Also, the Tyr183Met184Asp185Asp186 peptide is strictly conserved in all known immunodeficiency associated retroviral polymerases[1–5]. Compared with the unliganded RT structure, rotations of the side chains of Tyr181 and Tyr188 help to create a cavity to accommodate the NNRTIs during their entrance process[6]. Starting with the crystal structure conformation of the nevirapine-complexed RT, the ligand was omitted while the tertiary structure of RT was maintained. The segment Tyr181 to Tyr188 was excited, and acetyl and N-methylamino were introduced as N-terminal and C-terminal blocking groups, respectively (Fig. 1). Quite recently, it was shown [7–9] that this segment has a configuration which is like a Ω-loop which consists of two (not completely regular) chains that are connected by at least two hydrogen bonds.

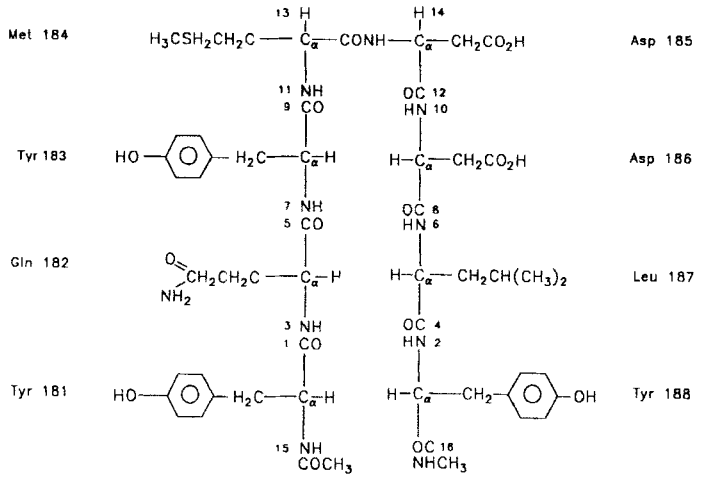

FIGURE 1 Amino acids of the Ω-loop. The numbers were used to code the intramolecular hydrogen-bonding forces.

Geometry Optimization of the Ω-Loop

X-ray crystallography may produce a considerable variability of bond lengths, bond angles, and torsion angles due to the experimental errors in X-ray determination. Also, the solid-state structure corresponds to one of perhaps many low-energy conformations. Optimization of crystalline-state geometry may lead to an improvement of geometry and the most stable conformer, therefore. The sequential steps of the combined molecular-mechanical and quantum-mechanical approach were as follows: geometry optimization with MM+ ⇒ QC to get atomic net charges ⇒ geometry optimization of the hybrid QC-MM+ structure ⇒ then geometry optimization with AMBER (isothermal-isobaric conditions at 0 kelvin temperature).

To perform explicit solvent simulations, the solvent was treated as a dielectric continuum surrounding the AMBER force field model of the solute molecules. A dielectric constants of $\varepsilon = 3.5$ was used which approximates a lipophilic environment of membrane-bound proteins, and of the blood-brain and gastrointestinal barrier[15].

Figure 2 shows the resulting Ω-loop. There are at least two strong O...H—N forces within the internal cage that maintain the Ω-loop. The first two amino acids that are involved in hydrogen bonding forces are Tyr181 and Tyr188 (H bond code 12 in Fig. 1). Figure 3 gives a schematic view.

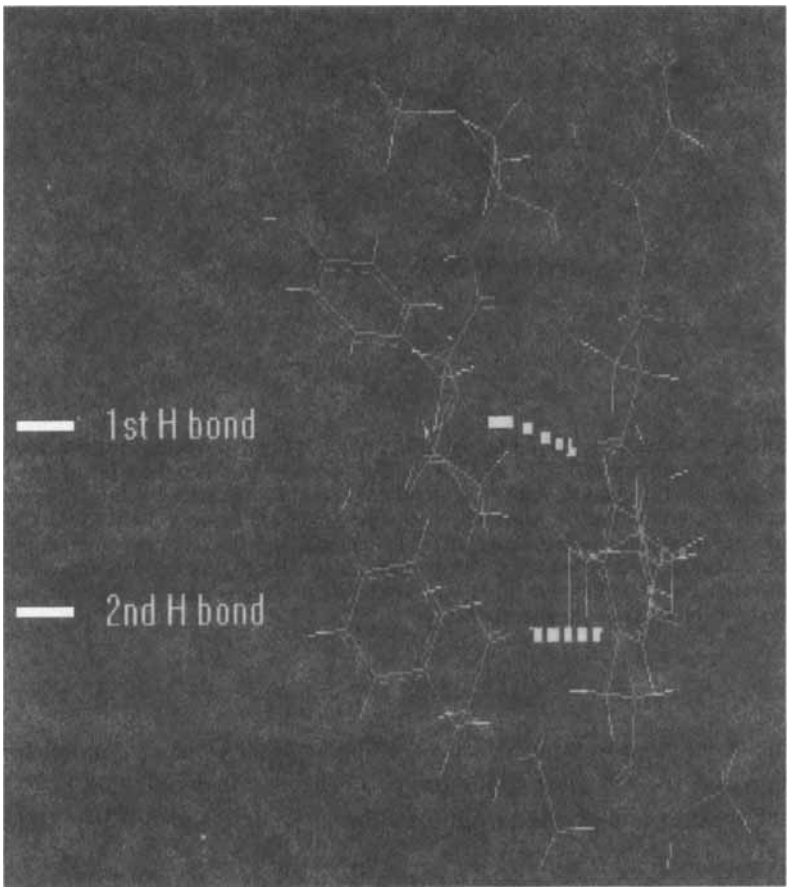

FIGURE 2 Geometry-optimized conformation of the Ω-loop. The configuration was obtained from nevirapine-complexed RT by X-ray analysis. After omission of the ligand, the segment Tyr181 to Tyr188 was excited, and acetyl and N-methylamino were introduced as N-terminal and C-terminal blocking groups. The geometry was optimized using a dielectric constant of $\varepsilon = 3.5$, 1 atm, and 0 Kelvin temperature. The two strongest intramolecular hydrogen bonds are indicated (see color plate I).

If the acetyl and N-methylamino groups of the N-terminal and C-terminal blocking moieties respectively, are replaced by the two naturally occuring amino acids, the sequences is as follows:

-[Val179-Ile180]-Tyr181-Gln182-Tyr183-Met184

-Asp185-Asp186-Leu187-Tyr188-[Val189-Gly190]-

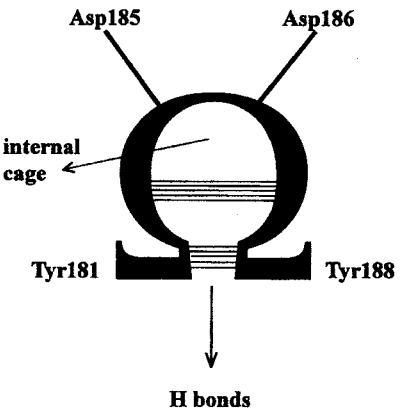

FIGURE 3 Schematic view of the Ω-loop with two "horizontal axes" that are formed by the nonnucleoside binding sites Tyr188 and Tyr188. The two aspartyl residues Asp185 and Asp186 of the catalytic triad (Asp110, Asp185, Asp186) are indicated. The strong hydrogen bonds of the peptide backbone that maintain the constrained loop are between Tyr181 and Tyr188 (independent on solvation effects), and within the internal cage (location and strength of this hydrogen bond depends on solvation effects and the used dielectric constant). Rotations of the side chains of Tyr181 and Tyr188 help to create a cavity to accommodate the NNRTIs during their entrance process into the pocket.

Setting A for an antiparallel β-sheet, P for a parallel β-sheet, and R for a random coil, the following configuration of this region is obtained by X-ray crystallography and geometry optimization if the unliganded RT is studied:

-[PA]-APPRRPPP-[PA]-

And for nevirapine-complexed RT, the following configuration is obtained:

-[RA]-AAPRRRPR-[PR]-

In both cases, the Ω-loop was maintained. The segment Tyr181 to Asp186 is highly hydrophilic[7], and water molecules are able to occupy the nonucleoside reverse transcriptase inhibitor (NNRTI) positions of the uncomplexed enzyme[17]. Prior nevirapine binding, desolvation is acquired therefore.

Geometry Optimization of Ordered Structures and the Random Coil

It is hypothesized that ordered configurations (for example, Ω-loop, α-helix, β-sheet, β-turn) are one of the conditions that peptides act as chemical

signal transducers by exciton and electron transfer[18, 19], while an unordered random coil blocks signal transfer through the molecular peptide network. It is generally accepted that noncompetitive inhibitors of the HIV-1 reverse transcriptase cause conformational changes of the Ω-loop[1–6].

On the other hand, the tendency of cascading β-sheets and α-helices to form highly anisotropic aggregates ("fibers") increases with increasing number of hydrogen-bonding forces, in dependence of the concentration of the peptide, a ligand such as a drug, solvent, *p*H, ionic strength, and temperature. Therefore, experimental studies have been suffering if, for example, the peptide concentration reaches a critical value, irrespective of the molecular weight[20] (β-sheet formation and subsequent aggregation are closely coupled events). This property to form insoluble aggregates renders studies[21–24] on the transition of β-sheets and α-helices to random coils *et vice versa* experimentally intractable, so that it was hoped that theoretical treatments and molecular simulations of model peptides provide useful informations [7–9, 21, 25–30].

Using the geometry-optimized Ω-loop as starting molecule, the angles of the backbone and residues were randomly varied. Then, reinitiation of an ordered structure was achieved by addition of ordered-structure-enforcing local constraints (torsional angles of the peptide backbone, defining the angles and distances of structure maintaining hydrogen bonds). The constraints were taken as a function of the degree of structural similarity of a number of known structures and conformations ("assignment rules"), as described in literature[31–34]. The standard geometry of each type of configuration is defined by the ϕ and ψ torsional angles of residues at the i and $(i+1)$ positions. Due to environment-dependent variations in amino acid residues, the sequence of amino acids, and hydrogen-bonding forces, the torsional angles vary statistically within the range of ± 30 degree[34] ("experimental error").

The most known helix is the (right-hand) α-helix ("alpha" in Table IV). The left-hand α-helix is called "left". 3_{10}-helices ("310") are less frequently found in proteins and also referred to be a special type of type I β-turns (the other subtypes of β-turns are here not considered). Parallel β-sheets are abbreviated as "parallel" in the table, and antiparallel β-sheets are called "beta". The double-C7 chair of the γ-turn is frequently referred as π-helix ("pi"). The extended γ-turn is abbreviated as "extend". A stochastic coil ("random") was designed by Monte Carlo technique. The resulting Cartesian coordinates are exemplified in Tables I and II with respect to the β-sheet and extended γ-turn. Figures 4 to 11 illustrate impressions on the geometry-optimized conformations. It is believed that the resulting geometry-optimized

TABLE I Example of Cartesian coordinates: begin and end of Cartesian coordinates of the antiparallel β-sheet conformation (SCHAKAL format)

TITL *CELL*	*<beta>*	*x*	*y*	*z*
ATOM	N1	−12.697	−3.804	3.740
ATOM	H1	−13.038	−4.730	3.522
ATOM	C1	−11.398	−3.900	4.383
ATOM	H2	−11.474	−3.170	5.191
ATOM	C2	−10.309	−3.374	3.437
ATOM	O1	−9.637	−4.177	2.788
ATOM	C3	−11.128	−5.297	4.982
ATOM	H3	−11.752	−5.412	5.869
ATOM	H4	−11.451	−6.096	4.313
ATOM	C4	−9.683	−5.551	5.437
ATOM	C5	−8.945	−6.680	4.990
ATOM	H5	−9.192	−7.193	4.071
ATOM	C6	−8.958	−4.654	6.268
ATOM	H6	−9.324	−4.411	7.254
ATOM	C7	−7.700	−6.840	5.639
ATOM	H7	−7.505	−7.657	6.318
ATOM	C8	−7.596	−4.474	5.913
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
ATOM	H61	9.635	7.498	−7.976
ATOM	C49	11.533	4.993	−7.724
ATOM	H62	11.379	4.243	−8.486
ATOM	C50	11.540	6.384	−7.978
ATOM	O15	12.635	7.053	−8.440
ATOM	H63	12.548	7.183	−9.386
ATOM	C51	10.373	3.223	−3.715
ATOM	O16	10.537	4.334	−3.168
ATOM	N10	11.327	2.310	−3.474
ATOM	C52	12.483	2.071	−2.849
ATOM	H64	12.992	1.088	−2.894
ATOM	H65	13.012	2.843	−2.255
ATOM	H66	11.060	1.462	−3.955
END				

conformations will be of sufficiently good quality to serve as valuable aids in solving the environmental-dependent transitions of conformations of biologically active peptide areas. At first glance, it can be visualized that the conformations of the peptide segment fall into a number of "structural classes".

Multivariate Correlation Between the Cartesian Coordinates of Molecules

There are two types of geometrical variables, the internal coordinate system which contains bond distances, bond angles, and dihedral angles; and the

TABLE II Example of Cartesian coordinates: begin and end of Cartesian coordinates of the extended γ-turn conformation (SCHAKAL format)

TITL CELL	*⟨extend⟩*	*x*	*y*	*z*
ATOM	N1	1.729	−7.159	−9.117
ATOM	H1	1.917	−8.075	−9.497
ATOM	C1	1.189	−7.292	−7.777
ATOM	H2	0.403	−6.539	−7.740
ATOM	C2	2.259	−6.896	−6.749
ATOM	O1	3.285	−7.570	−6.642
ATOM	C3	0.638	−8.711	−7.545
ATOM	H3	−0.182	−8.835	−7.252
ATOM	H4	1.375	−9.487	−7.753
ATOM	C4	0.042	−8.881	−6.143
ATOM	C5	0.751	−9.498	−5.085
ATOM	H5	0.809	−10.569	−4.963
ATOM	C6	−1.086	−8.156	−5.682
ATOM	H6	−2.101	−8.453	−5.904
ATOM	C7	1.106	−8.570	−4.077
ATOM	H7	2.128	−8.291	−3.867
ATOM	C8	−0.732	−7.226	−4.675
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
ATOM	H61	−7.205	8.546	6.998
ATOM	C49	−6.162	9.618	4.062
ATOM	H62	−6.343	10.002	3.069
ATOM	C50	−7.114	9.690	5.114
ATOM	O15	−7.899	10.788	5.317
ATOM	H63	−8.717	10.692	4.824
ATOM	C51	−3.320	5.958	5.793
ATOM	O16	−4.229	5.201	6.190
ATOM	N10	−2.542	6.498	6.739
ATOM	C52	−1.511	7.317	6.962
ATOM	H64	−1.119	7.524	7.979
ATOM	H65	−0.970	7.833	6.142
ATOM	H66	−2.853	6.152	7.636
END				

(x, y, z)-matrix of Cartesian coordinates. Using the two methods there are certain advantages and limitations with respect to unconstrained and constrained geometry optimization, and analytic solvent-accessible surface calculations of molecular systems [35–37]. However, the coordinates of the final geometry are exchangeable by algorithms which are implemented in molecular modelling software.

To compare two sets of molecules (e.g., ligands in receptor sites, agonist and antagonist, various antagonists of a series), Cartesian coordinates were used because the internal coordinates are scale sensitive[38]. The multivariate

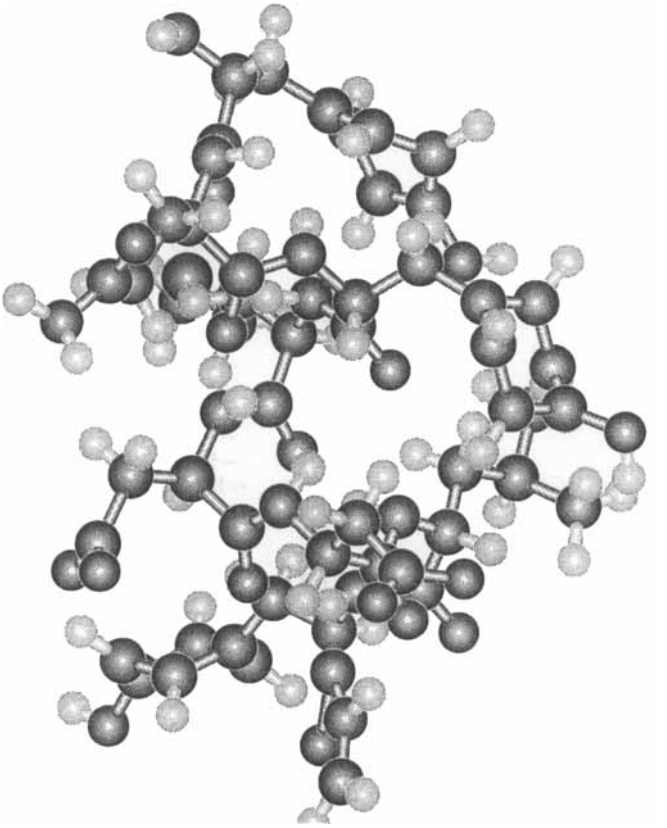

FIGURE 4 Geometry-optimized conformation of the 3_{10}-helix ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

TABLE III Example of canonical analysis: ordered eigenvalues ($f = 1, 2, 3$], canonical correlation coefficients, and canonical variates of the multivariate correlation between the first subset of Cartesian coordinates (antiparallel β-sheet conformation, Table I) and the second subset (extended γ-turn conformation, Table II).

Notation		*f = 1*	*f = 2*	*f = 3*
Eigenvalues		0.950	0.293	0.008
Correlation		0.975*	0.542	0.281
Variates				
Subset 1	*x*	−0.103	−0.881	−0.871
	y	−0.335	1.257	−1.933
	z	0.628	0.660	−2.383
Subset 2	*x*	0.175	0.132	1.301
	y	−1.084	1.001	0.809
	z	0.301	−1.460	0.247

*See "extend-beta" of Table IV.

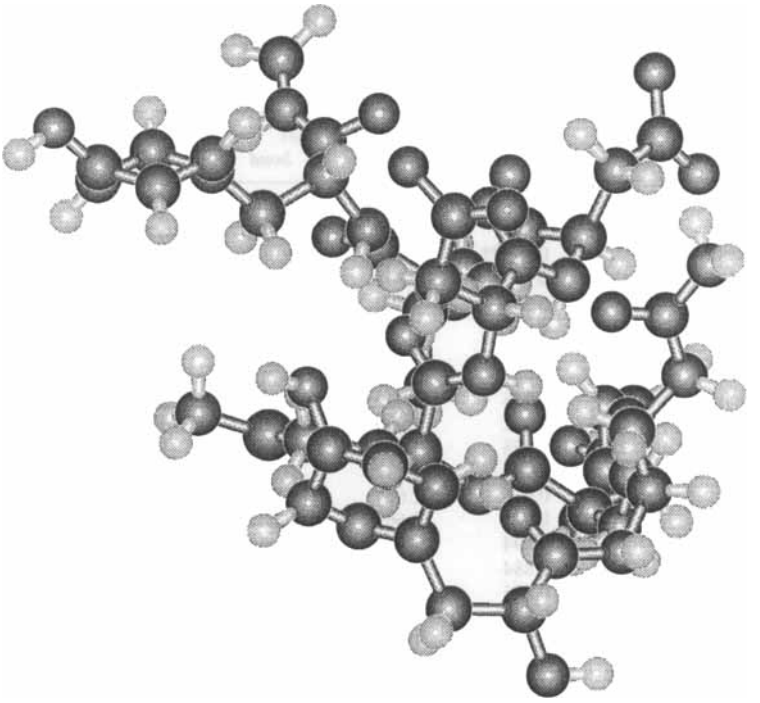

FIGURE 5 Conformation of the left-hand α-helix ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

and canonical correlation module of the MASCA model[39,40] is applied in order to get a statistically testable criterion on the similarity of subsets of atomic points of two molecules[41]. Details of canonical correlation see literature[40,42–46].

Let us return to the two Cartesian matrices of Tables I and II. The results of a canonical correlation analysis are summarized in Table III. The two eigenvalues are significantly different from zero at the 5% level or less. Figure 12 shows the first canonical plot. The straight line led to a correlation coefficient of $\rho = 0.975$ of the two first canonical variates (that are subsets of the Cartesian coordinates). Therefore, the two conformations of the peptide segment do not differ markedly. The other canonical correlation coefficients are listed in Table IV.

Another, more intuitive way is to determine the root mean square (*RMS*) of the two conformations to be compared in such a way that the *RMS* distance between corresponding atoms is minimized. This implies: the lower the *RMS* value, the greater the conformational similarity. The *RMS*s of the

TABLE IV Comparison of conformational families: canonical correlation coefficient R_{can} between the corresponding atoms in a pair of conformations, and root mean squares (*RMS*s) of the internal coordinates (bonds, bond angles, torsional angles) and Cartesian coordinates (Cartes.)

Conformations	R_{can}	*RMS values* Cartes.	*bond*	*angle*	*tors.*
alpha-loop	1.000*	0.153	0.001	1.121	9.911
extend-pi	0.998	1.297	0.002	1.677	53.783
extend-beta	0.975**	1.415	0.002	2.282	45.893
beta-pi	0.975	1.478	0.003	2.699	44.501
310-extend	0.966	2.399	0.687	18.620	52.976
extend-parall	0.960	1.789	0.002	1.739	53.405
parall-pi	0.958	1.467	0.001	1.616	48.440
310-left	0.951	2.686	0.710	19.558	54.206
extend-left	0.948	2.039	0.004	2.717	56.848
left-pi	0.942	1.990	0.004	2.976	51.627
310-pi	0.938	1.659	0.654	18.444	55.434
parall-beta	0.935	1.237	0.003	2.351	40.202
310-parall	0.933	1.604	0.705	19.109	56.544
left-parall	0.901	2.663	0.004	2.495	46.428
extend-random	0.894	2.400	0.002	2.209	54.546
310-random	0.884	3.306	0.652	18.277	48.728
310-beta	0.876	1.904	1.031	23.035	58.125
pi-random	0.872	2.439	0.003	2.353	57.784
beta-random	0.864	1.352	0.002	1.352	57.689
left-beta	0.851	2.593	0.005	2.112	52.284
left-random	0.840	2.725	0.004	2.073	57.565
alpha-extend	0.816	2.905	2.644	30.286	30.286
extend-loop	0.816	2.912	1.621	31.371	51.961
310-alpha	0.804	2.684	1.594	30.295	51.053
310-loop	0.801	2.689	1.582	30.129	51.324
alpha-parall	0.783	2.893	2.768	29.967	54.447
alpha-pi	0.783	2.764	0.002	1.677	51.092
loop-parall	0.783	2.901	2.768	30.032	54.426
loop-pi	0.782	2.772	2.504	30.003	50.981
parall-random	0.772	2.593	0.003	2.038	57.564
alpha-left	0.761	2.785	2.191	30.535	52.868
left-loop	0.758	2.812	1.621	31.183	51.905
alpha-beta	0.751	3.469	1.631	31.315	50.169
loop-beta	0.748	3.473	1.621	31.017	50.171
loop-random	0.735	2.530	2.021	30.211	51.542
alpha-random	0.734	2.530	0.002	2.209	51.599

* 0.99998.

** See Table III, footnote.

internal coordinates[47] are given in Table IV. The squared canonical correlation coefficient may be approximated by:

$$R^2_{can} = 1.070 - 0.122\ RMS(Cart.) - 0.044\ RMS(bond)$$

where the squared multiple correlation coefficient $R^2 = 0.631$ and the two regression coefficient are significant at the 5% level or less (the other two

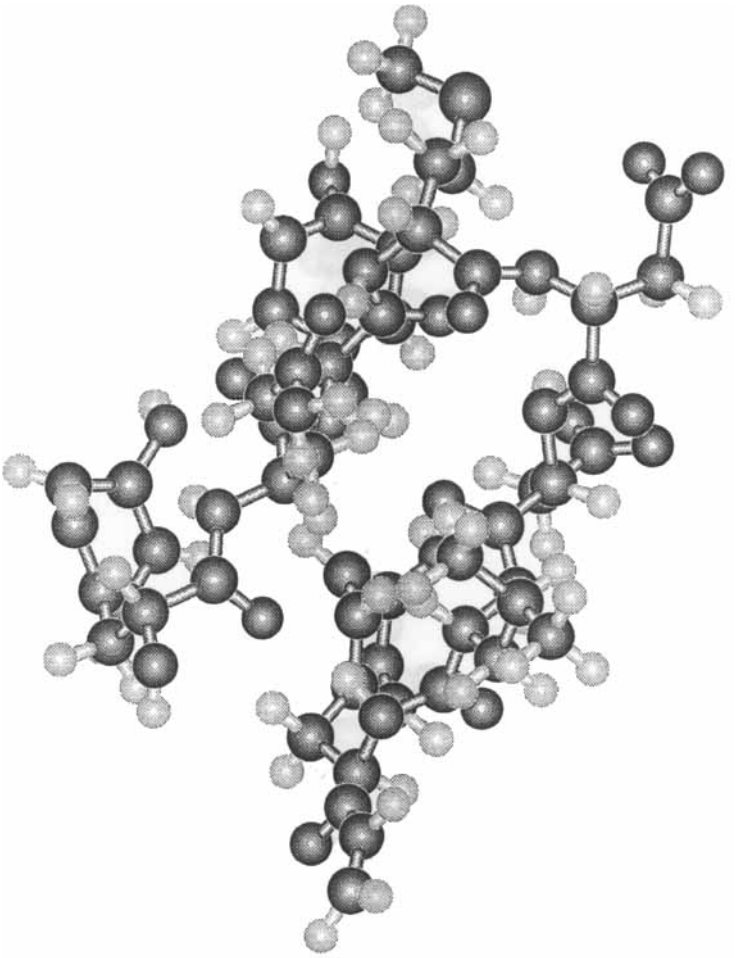

FIGURE 6 Conformation of the right-hand α-helix ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

RMS parameters were insignificant). Figure 13 shows the plot. Therefore, the best way in comparing quantitatively conformations is to estimate directly the canonical correlation coefficient.

The results show (Table IV) that there is a hierarchy of subgroups of geometrically similar conformers. The similarity is great between the Ω-loop and α-helix, extended γ-turn and π-helix, and extended γ-turn and β-sheet. A good clustering becomes intuitively clear if two conformations are overlayed on a computer screen. For example, Figure 14 shows the superim-

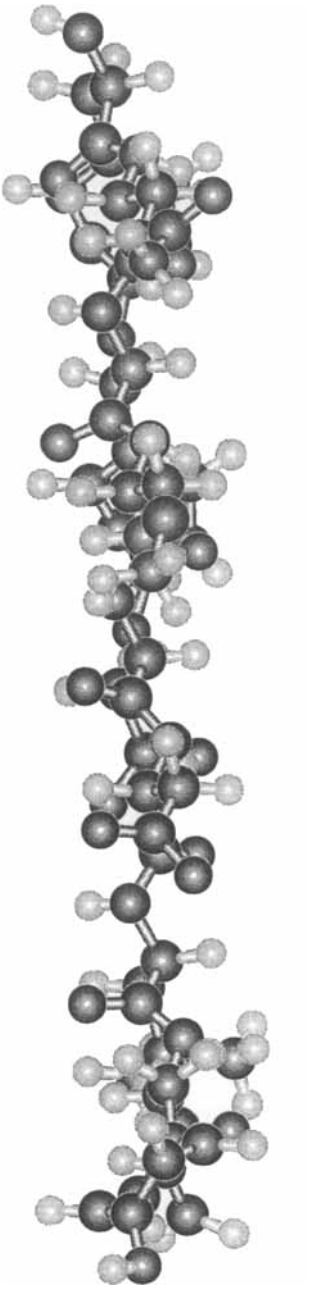

FIGURE 7 Conformation of the antiparallel β-sheet ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

FIGURE 8 Conformation of the parallel β-sheet ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

posed extended γ-turn and β-sheet ($R_{can} = 0.975$, see Table IV).The lowest similarity is found with respect to the Ω-loop and the random coil, and the α-helix and random coil.

Cluster and Factor Analysis of Multiconformational Structures

The usefulness of considering ensemble of conformers, rather than identifying low energy minima, by cluster analysis, was demonstrated quite recently [48–50]. The following physicochemical descriptors were used in literature: torsional angles[48,49], energies[48], atomic distances[48], and subsets of atoms[50]. In this example, the variable is the canonical correlation coefficient (Table IV). In a subsequent cluster analysis, it is required that

FIGURE 9 Conformation of the extended α-sheet ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

clusters are formed by fusion. Therefore, the agglomerative method is applied to the resulting (9,9)-matrix of canonical correlation coefficients (Ward linkage). Figure 15 shows the result. The α-helix and Ω-loop ("cases" 1 and 2) are closely related in conformation so that they form the first cluster. Then, the π-helix and extended γ-turn ("cases" 5 and 3), the 3_{10}-helix and left-hand α-helix ("cases" 6 and 8), and the antiparallel and parallel β-sheets ("cases" 4 and 7) are conformationally related. These configurations form the second cluster. The random coil ("case" 9) forms an own cluster.

To examine the cluster results by an additional technique, factor analysis (orthomax rotation) was applied (Fig. 16). Also, the α-helix and Ω-loop

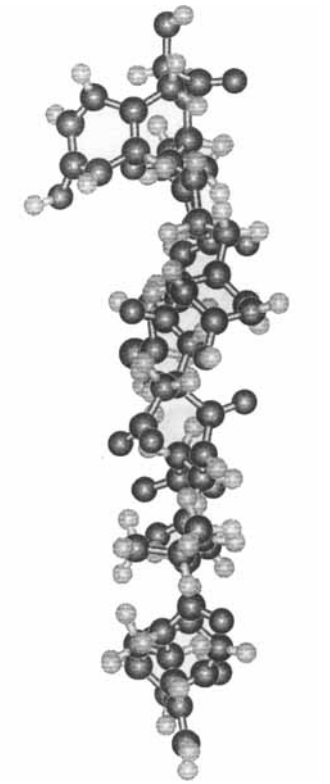

FIGURE 10 Conformation of the double-C7 γ-turn ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

("var" 1 and 2) are closely related. As expected, the random-coil ("var" 9) is completely discriminated from the remaining subset of conformations.

4. DISCUSSION AND CONCLUSIONS

The segment Tyr181 to Tyr188 of the HIV-1 reverse transcriptase molecule contains two amino acids (Asp185, Asp186) of the catalytic aspartyl triad (Asp110, Asp185, Asp186) and two amino acids (Tyr181, Tyr188) of the nonnucleoside RT inhibitor (NNRTI) binding sites. In order to investigate its structure-specifying forces, it was dissected from the HIV-1 RT. The folding patterns are based on at least two intramolecular hydrogen bonds

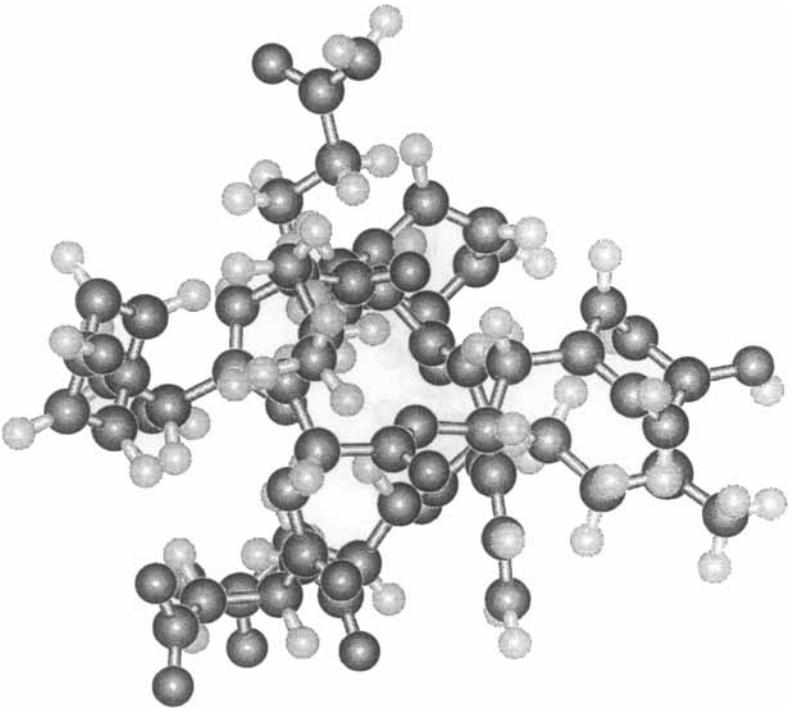

FIGURE 11 Conformation of the random coil ($\varepsilon = 3.5$, 1 atm, 0 Kelvin temperature).

which hold two peptide chains together so that a constrained Ω-loop configuration is found (Figs. 2 and 3).

Geometric constraints were introduced, particularly with respect to the backbone and side-chain conformations and their hydrogen bonds, to simulate the minimum-energy conformations of a random coil, and of ordered structures (Ω-loop, 3_{10}-helix, left-hand and right-hand α-helix, antiparallel and parallel β-sheet, extended β-sheet, double-C7 γ-turn).

To compare statistically the conformations, Cartesian coordinates were determined from internal coordinates. As criterion of conformational similarity, the canonical correlation coefficients of two molecular subsets of Coordinates were estimated (Table IV). There is a hierarchy of clustering of geometrically similar conformers. For example, the similarity is large between the Ω-loop and α-helix, extended γ-turn and π-helix, and extended γ-turn and β-sheet, and lower for the Ω-loop and the random coil, and the α-helix and random coil. This good clustering becomes intuitively clear if

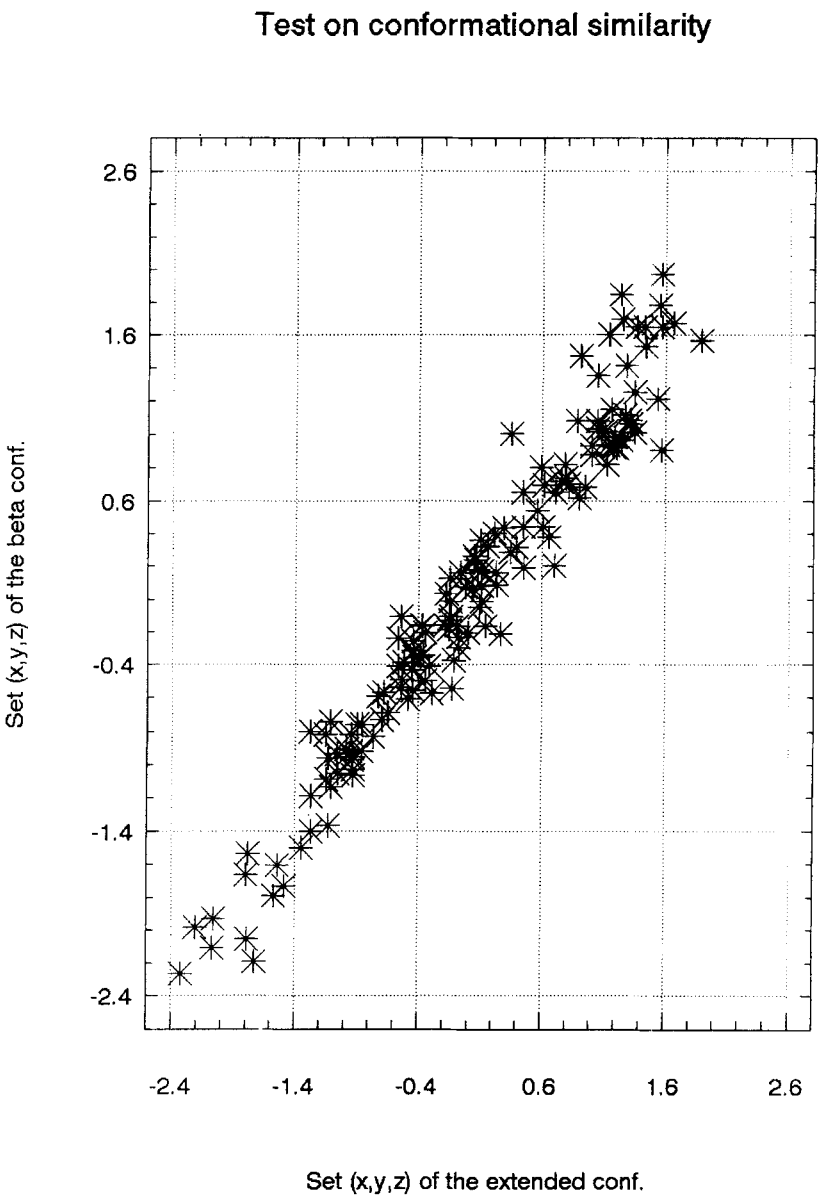

FIGURE 12 Plot of the scores of canonical variates derived from the Cartesian coordinates of the extended and antiparallel β-sheets (return to Table III, see also Fig. 14).

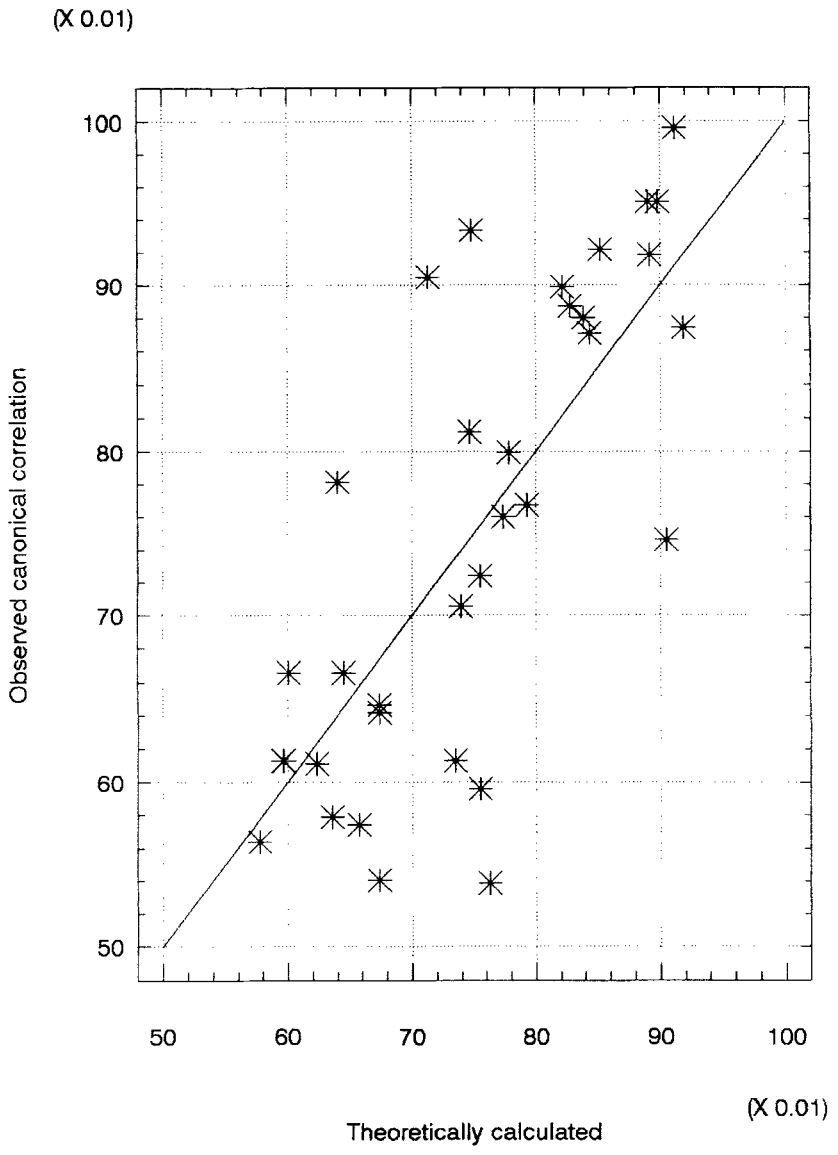

FIGURE 13 Plot of the squared canonical correlation coefficients *vs.* a linear combination of *RMS*s of bonds and Cartesian coordinates (return to Table IV).

two conformations are superimposed (for example, see Fig. 14). The use of cluster and factor analysis gave deeper insight. The α-helix and Ω-loop form a first cluster. The π-helix and extended γ-turn, the 3_{10}-helix and left-hand α-helix, and the antiparallel and parallel β-sheets form a second cluster. The

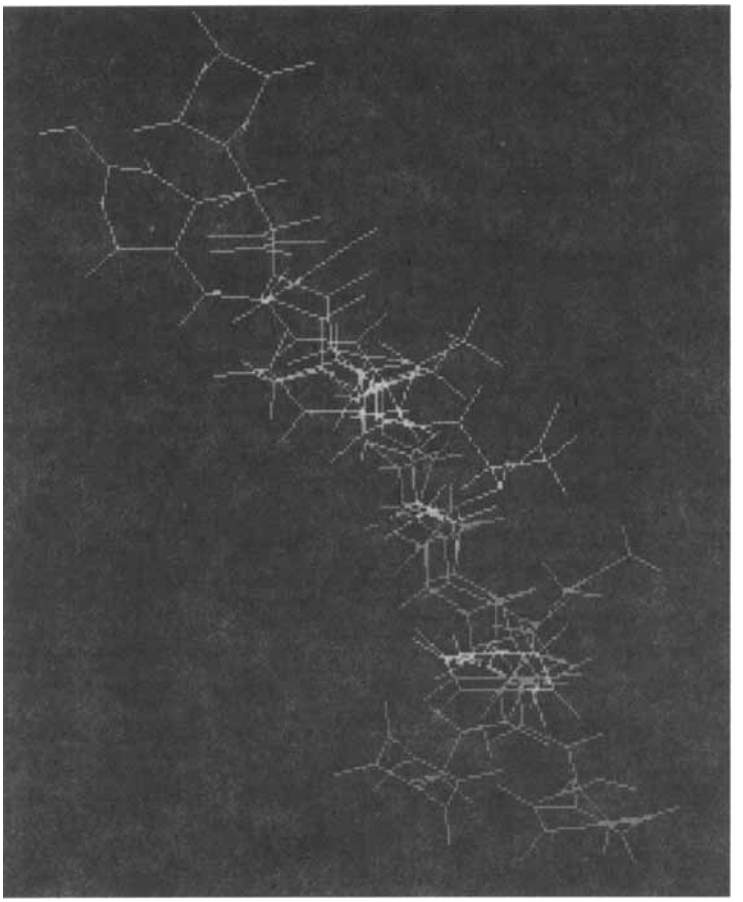

FIGURE 14 Superposition of the extended (white color) and antiparallel β-sheets (yellow). Return to Table 3 and Figure 12, and Figures 7 and 9 (see color plate II).

random coil forms a third cluster and is completely discriminated from the remaining subset of conformations (Figs. 15 and 16).

The results may lead to the following speculations: It is generally accepted that noncompetitive inhibitors of the HIV-1 reverse transcriptase cause conformational changes of the naturally occuring Ω-loop. The study suggests that the Ω-loop may be converted after ligand binding to another conformation, probably by changing the dynamic equilibrium between helical, β-sheet, γ-turn, and even random coil conformations. The type of the transition of the dynamic equilibria depends on the type and number of

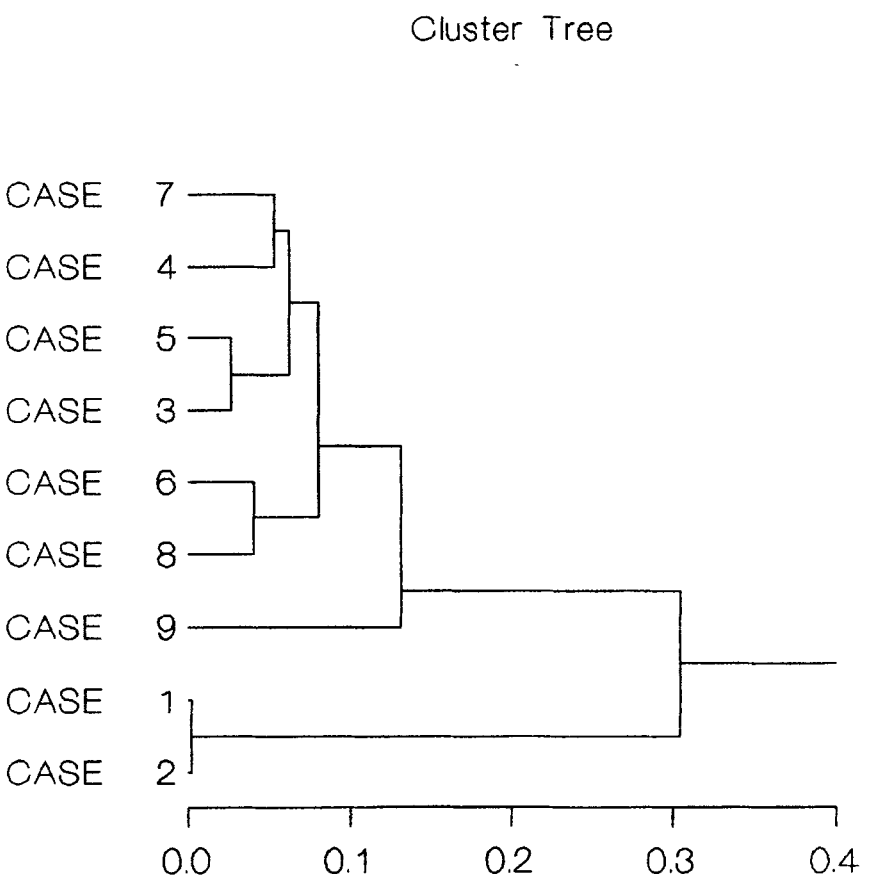

FIGURE 15 Results of cluster analysis. Notations and explanations see text.

amino acids involved in binding, that is, it depends on the structure of the drugs bound onto the amino acid residues. If allosteric interactions play a role, the inhibitors could stabilize the induced conformations, in particular, the transition to a random coil of the Tyr181 to Tyr188 segment. It might be expected that this conformation "produces" an inactive enzyme. (It should be noted that the random coil conformation of the Aβ peptide, the major protein constituent of senile plaques in patients with Alzheimer's disease, are poorly toxic or inactive, while the β-sheet conformation causes Alzheimer's disease[24]).

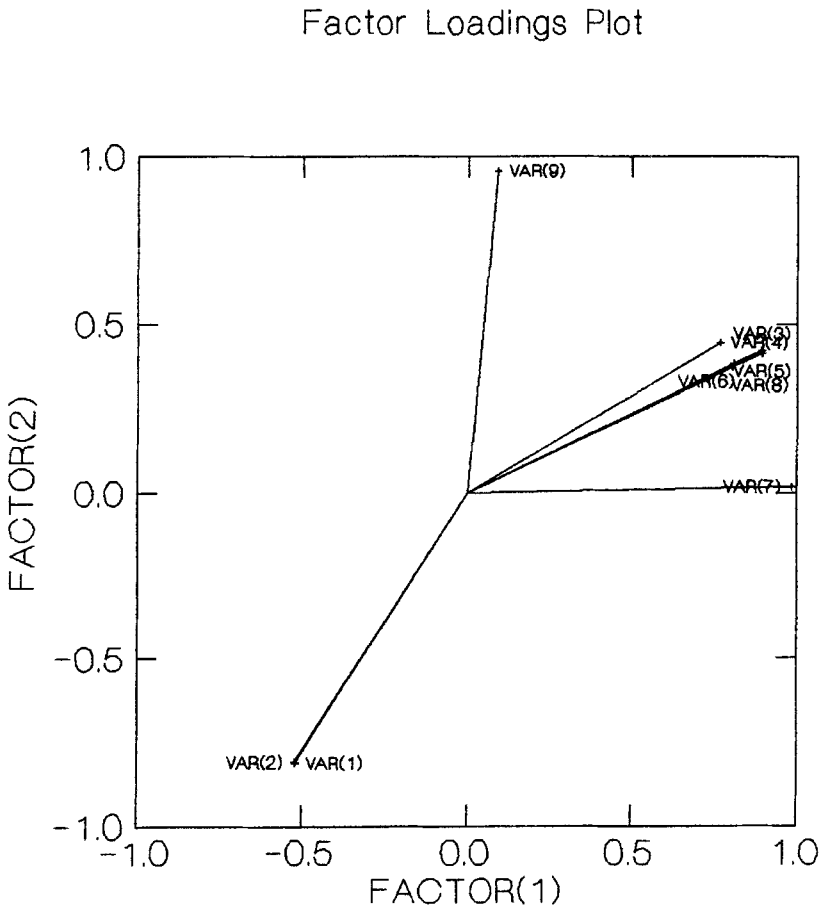

FIGURE 16 Results of factor analysis. Notations and explanations see text.

References

[1] Ren, J. S., Esnouf, R., Garman, E., Somers, D., Ross, O., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. and Stammers, D. (1995). High Resolution Structures of HIV-1 RT from Four RT-inhibitor Complexes. *Nature Struct. Biol.*, **2**, 293–302

[2] Ding, J., Das, K., Tantillo, C., Zhang, W., Clark, A. D., Jessen, S., Lu, X., Hsiou, Y., Jacobomolina, A., Andries, K., Pauwels, R., Moereels, H., Koymans, L., Janssen, P. A. J., Smith, R. H., Koepke, M. K., Michejda, C. J., Hughes, S. H. and Arnold, E. (1995). Structure of HIV-1 Reverse Transcriptase in a Complex with the Non-nucleoside Inhibitor Alpha-APA R 95845 at 2.8 Angstrom Resolution. *Structure*, **3**, 365–379.

[3] Boyer, P. L., Ferris, A. L., Clark, P., Whitmer, J., Frank, P., Tantillo, C., Arnold, E. and Hughes, S. H. (1994). Mutational Analysis of the Fingers and Palm subdomains of Human Immunodeficiency Virus Type-1 (HIV-1) Reverse Transcriptase. *J. Mol. Biol.*, **243**, 472–483.

[4] Pauwels, R., Andries, K., Janssen, P. A. J. and Arnold, E. (1994). Locations of Anti-AIDS Drug Binding Sites and Resistance Mutations in the Three-Dimentional Structure of HIV-1 Reverse Transcriptase – Implications for Mechanisms of Drug Inhibition and Resistance. *J. Mol. Biol.*, **243**, 369–387.

[5] Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A. J. and Arnold, E. (1994). Locations of Anti-AIDS Drug Binding Sites and Resistance Mutations in the Three-dimensional Structure of HIV-1 Reverse Transcriptase. *J. Mol. Biol.*, **243**, 369–387.

[6] Das, K., Ding, J., Hsiou, Clark, Jr., A. D., Zhang, W., Moereels, H., Koymans, L., Andries, K., Janssen, P. A. J., Hughes, S. H. and Arnold, E. (1996). A Comparison of the Structures of HIV-1 RT Complexes with Two TIBO Derivatives, of which One is a Potential Drug for AIDS. *Abstr. Symp. Structure-Based Drug Design, June, 13–16, Rutgers Univ., NJ*, in: *Drug Design & Discovery*, **13**, 159.

[7] Mager, P. P. and Walther, H. (1996). A Hydrophilic Ω-Loop (Tyr181 to Tyr188) in the Nonsubstrate Binding Area of HIV-1 Reverse Transcriptase. *Drug Design & Discovery*, in press.

[8] Mager, P. P. (1996). Molecular Simulation of the Folding Patterns of the Ω-Loop (Tyr181 to Tyr188) of HIV-1 Reverse Transcriptase. *Drug Design & Discovery*, in press.

[9] Mager, P. P. Temperature-dependent Changes in the Folding Pattern of the Ω-Loop (Tyr181 to Tyr188) of HIV-1 Reverse Transcriptase. *Drug Design & Discovery*, in press.

[10] Allinger, N. L. and Yan, L. (1993). Molecular Mechanics (MM3). Calculation of Furan, Vinyl, Ethers, and Related Compounds. *J. Am. Chem. Soc.*, **115**, 11918–11925.

[11] Tai, J. C., Yang, L. and Allinger, N. L. (1993). Molecular Mechanics (MM3). Calculation on Nitrogen-Containing Aromatic Heterocyclics. *J. Am. Chem. Soc.*, **115**, 11906–11917.

[12] Weiner, S. J., Kollman, P. A., Nguyen, D. T. and Case, D. A. (1986). An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.*, **7**, 230–252.

[13] Gasteiger, J. and Marsili, M. (1980). Iterative Partial Equalization of Orbital Electronegativity – a Rapid Access to Atomic Charges. *Tetrahedron*, **36**, 3219–3288.

[14] Frey, R. F., Coffin, J., Newton, S. Q., Ramek, M., Cheng, V. K. W., Momany, F. A. and Schäfer, L. (1992). Importance of Correlation-Gradient Geometry Optimization for Molecular Conformational Analysis. *J. Am. Chem. Soc.*, **114**, 5369–5377.

[15] Mager, P. P. (1994). Interactive Multivariate Modelling of ArgGlyAsp (RGD) Derivatives. *Med. Res. Rev.* (New York), **14**, 75–126.

[16] Brookhaven Protein Data Bank (PDB), files 3HVT. ENT and 1RTI. ENT.

[17] Hsiou, Y., Ding, J., Das, K., Clark, Jr. A. D., Hughes, S. H. and Arnold, E. (1996). Crystal Structure of Unliganded HIV-1 Reverse Transcriptase. *Abstr. Symp. Structure-Based Drug Design, June 13–16, 1995, Rutgers Univ., NJ*, in: *Drug Design & Discovery*, **13**, 158.

[18] Mager, P. P. (1997). Some Remarks on Elementary Processes in Quantum Cell Chemistry. *Acta Histochem.*, **53**, 302–305.

[19] Gretchikhine, A. B. and Ogawa, M. Y. (1996). Photoinduced Electron Transfer Along a β-Sheet Mimic. *J. Am. Chem. Soc.*, **118**, 1543–1544.

[20] Maeda, H., Gatto, Y. and Ikeda, S. (1984). Effects of Chain Length and Concentration of the β-Coil Conversion of Poly [S-carboxymethyl)-L-cysteine] in 50 nM NaCl solution. *Macromolecules*, **17**, 2031–2037.

[21] Kanô, F. and Maeda, H. (1996). Monte Carlo Simulation of the β-Sheet-Random Coil Transition of a Homopolypeptide. I. Equilibrium Study. *Mol. Simul.*, **16**, 261–274.

[22] Kemp, D. S., Allen, T. J. and Ostlick, S. L. (1995). The Energetics of Helix Formation by short-Templated Peptides in Aqueous Solution. 1. Characterization of the Reporting Helical Template Ac-Hel_1. *J. Am. Chem. Soc.*, **117**, 6641–6657.

[23] Perderson, D., Gabriel, D. and Hermans, J. Jr., (1971). Potentiometric Titration of Poly-L-lysins: the Coil-to-β Transition. *Biopolymers*, **10**, 2133–2145.

[24] Simmonis, L. K., May, P. C., Tomaselli, K. J., Rydel, R. E., Fuson, K. S., Brigham, E. F., Wright, S., Lieberburg, I., Becker, G. W., Brems, D. N. and Li, W. Y. (1994). Secondary Structure of Amyloid β Peptide Correlates with Neurotoxic Activity *In Vitro*. *Mol. Pharmacol.*, **45**, 373–379.

[25] Kanô, F. (1976). Theory of the phase Transition Between the Intra β-Structure and the Random Coil in Polyamino Chains. *J. Phys. Soc. Jp.*, **41**, 219–227.

[26] Wakana, H., Shigaki, T. and Saitô, N. (1982). Intramolecular α-Helix/β-Structure/Random Coil Transition in Polypeptides. I. Equilibrium Case. *Biophys. Chem.*, **16**, 275–285.
[27] Mattice, W. L. and Scheraga, H. A. (1984). Matrix Formulation of the Transition from a Statistical Coil to an Intramolecular Antiparallel β Sheet. *Biopolymers*, **24**, 1701–1724.
[28] Mattice, W. L. and Scheraga, H. A. (1985). Role of Interstand Loops in the Formation of Intramolecular Cross-β-Sheets by Homopolyamino Acids. *Biopolymers*, **24**, 565–579.
[29] Mattice, W. L. (1989). The β-Sheet to Coil Transition. *Annu. Rev. Biophys. Biochem.*, **18**, 93–111.
[30] Carlacci, L. and Englander, S. W. (1996). Loop Problem in Proteins: Developments on the Monte Carlo Simulated Annealing Approach. *J. Comput. Chem.*, **17**, 1002–1012.
[31] Venkatachalam, C. M. (1968). Stereochemical Criteria for Polypeptides and Proteins. V. Conformation of a System of Three Linked Peptide Units. *Biopolymers*, **6**, 1425–1436.
[32] Perczel, A., McAllister, M. A., Császár, P. and Czizmadia, I. G. (1993). Peptide Models. 6. New β-Turn Conformations from *ab initio* Calculations Confirmed by X-Ray Data of Proteins. *J. Am. Chem. Soc.*, **155**, 4849–4858.
[33] Böhm, H. J. (1993). *Ab intio* SCF Calculations on Low-Energy Conformers of N-Acetylglycine N′-Methylamide. *J. Am. Chem. Soc.*, **115**, 6152–6158.
[34] Xie, P., Zhou, O. and Diem, M. (1995). Conformational Studies on β-Turns in Cyclic Peptides by Vibrational CD. *J. Am. Chem. Soc.*, **117**, 9502–9508.
[35] Baker, J. and Bergeron, D. (1993). Constrained Optimization in Cartesian Coordinates. *J. Comput. Chem.*, **14**, 1339–1346.
[36] Alsberg, B. K., Jensen, V. R. and Børve, K. J. (1996). Use of Multivariate Methods in the Analysis of Calculated Reaction Pathways. *J. Comput. Chem.*, **17**, 1197–1216.
[37] Eisenhaber, F. and Argos, P. (1993). Improved Strategy in Analytic Surface Calculation for Molecular Systems: Handling of Singularities and Computational Efficiency. *J. Comput. Chem.*, **14**, 1272–1280.
[38] Susnow, R., Schutt, C. and Rabitz, H. (1994). Principal Component Analysis of Dipeptides. *J. Comput. Chem.*, **15**, 963–980.
[39] Mager, P. P. (1975). The MASCA Model of Biochemical Pharmacological Drug Research. *Drug. Res.*, **25**, 1006–1008, 1270–1272, 1355–1356, 1475–1476, 1745, 1864–1865.
[40] Mager, P. P. (1991). *Design Statistics in Pharmacochemistry*. New York: Wiley.
[41] Mager, P. P. (1992). Multivariate Analysis of Cartesian Coordinates of Bioorganic Molecules. In *QSAR in Design of Bioactive Compounds*, edited by Kuchar, M. pp. 446–469. Barcelona, J. R. Prous Sci. Publ.
[42] Rencher, A. C. (1988). On the Use of Correlations to Interpret Canonical Functions. *Biometrika*, **75**, 363–365.
[43] Flury, B. W. (1989). Understanding Partial Statistics and Redundancy of Variables in Regression and Discriminant Analysis. *Am. Stat.*, **43**, 27–31.
[44] Krzanowski, W. J. (1989). On Confidence Regions in Canonical Variate Analysis. *Biometrika*, **76**, 107–116.
[45] Rencher, A. C. (1992). Interpretation of Canonical Discriminant Functions, Canonical Variates, and Principal Components. *Am. Stat.*, **46**, 217–225.
[46] Neuenschwander, B. E. and Flury, B. D. (1995). Common Canonical Variates. *Biometrika*, **82**, 553–560.
[47] Treasurywala, A. M., Jaeger, E. P. and Peterson, M. L. (1996). Conformational Searching Methods for Small Molecules. III. Study of Stochastic Methods Available in SYBYL and MACROMODEl. *J. Comput. Chem.*, **17**, 1171–1182 (and refs. 5–7 quoted therein).
[48] Michel, A. G. and Jeandenans, C. (1993). Multiconformational Investigations of Polypeptide Structures, Using Clustering Methods and Principal Components Analysis. *Computers Chem.*, **17**, 49–59.
[49] Shenkin, P. S. and McDonald, D. Q. (1994). Cluster Analysis of Molecular Conformations. *J. Comput. Chem.*, **15**, 899–916.
[50] Torda, A. E. and van Gunsteren, W. F. (1994). Algorithms for Clustering Molecular Dynamics Configurations. *J. Comput. Chem.*, **12**, 1331–1340.