Theoret. chim. Acta (Berl.) 22, 11—22 (1971)

Molecular Orbital Calculations on the Conformation of Polypeptides and Proteins

IX. The Conformational Energy Maps of the Lysyl and Arginyl Residues*

B. PULLMAN, J. L. COUBEILS, P. COURRIERE, and D. PERAHIA
Université de Paris, Institut de Biologie Physico-Chimique, 13 rue P. et M. Curie, Paris 5è

Received December 22, 1970

The conformational energy maps are computed for the lysyl and the arginyl residues with the help of the quantum-mechanical method PCILO. Because of the relatively very large number of the possible combinations of the side chain rotational angles, sub-maps are constructed only for the most common such combinations as indicated from the X-rays study of ten globular proteins. These sub-maps are then combined for the construction of the general conformational energy map. The comparison of the theoretically allowed and experimentally observed values of backbone dihedral angles Φ and Ψ for the lysyl and arginyl residues in the globular proteins indicates a very satisfactory agreement among the two. These results confirm that the contour of the stable zone on the conformational energy map depends primarily on χ^1 but indicate also that its fine structure is sensitive to the remaining χ's in particular to χ^2.

Diagramme der Konformationsenergie werden für den Lysyl- sowie den Arginylrest mit Hilfe der quantenmechanischen PCILO-Methode berechnet. Wegen der großen Zahl möglicher Kombinationen der Rotationswinkel der Seitenketten werden Teildiagramme nur für die wichtigsten derartigen Kombinationen, die bei Untersuchungen nach der Röntgenstrahlmethode an zehn globulären Proteinen gefunden wurden, konstruiert. Diese Teildiagramme werden dann zum Gesamtdiagramm der Konformationsenergie kombiniert. Die Übereinstimmung zwischen experimentellen und theoretisch erlaubten Winkeln Φ und Ψ der Hauptkette und den Lysyl- und Arginylresten ist sehr zufriedenstellend. Diese Ergebnisse bestätigen, daß der stabile Bereich des Diagramms der Konformationsenergie hauptsächlich von χ^1 abhängt; sie zeigen aber auch, daß seine genaue Struktur von den übrigen χ-Werten beeinflußt wird, insbesondere von χ^2.

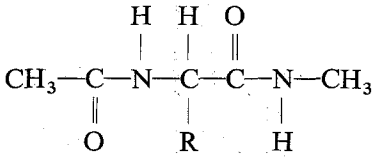
Les cartes d'énergie conformationelle des résidus lysyle et arginyle sont construites à l'aide de la méthode PCILO de la Chimie Quantique. A cause du nombre très élevé des combinaisons possibles des angles rotationnels de la chaîne latérale, des sous-cartes sont construites pour les plus fréquentes seulement de ces combinaisons (selon les indications de l'étude aux rayons X de dix protéines globulaires). Ces sous-cartes sont ensuite combinées pour la construction de la carte générale d'énergie conformationnelle. La comparaison des valeurs théoriques et expérimentales des angles Φ et Ψ de la chaîne principale pour les résidus lysyle et arginyle des protéines indique un accord très satisfaisant entre les deux. Ces résultats confirment que le contour de la zone stable sur la carte d'énergie conformationnelle est déterminé principalement par la valeur de χ_1 mais indiquent aussi que la structure fine de cette zone est sensible aux valeurs des χ supérieurs, en particulier de χ^2.

Introduction

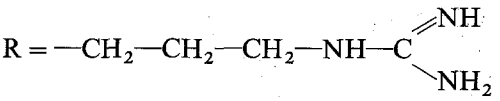
Continuing our quantum mechanical studies on the conformational energy maps of the amino-acid residues of proteins [1–8], we have extended the computations to the lysyl and arginyl residues.

* This work was supported by grant n° 67-00-532 of the Délégation Générale à la Recherche Scientifique et Technique (Comité de Biologie Moléculaire).

The calculations have been performed on the model "dipeptides":

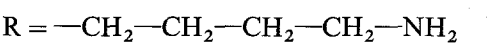


with



for the arginyl residue

and



for the lysyl residue.

We have followed, as previously, the general conventions for the description of the conformations as indicated in [9] with the notations represented in Figs. 1 and 2. The method utilized is, also as in the previous papers of this series, the PCILO method (Perturbative Configuration Interaction using Localized Orbitals) [10]. The geometrical input data (bond lengths and bond angles) for the residues have been taken from the compilation of Gurskaia [11].

One distinct difference was, however, applied in the procedure utilized for the lysyl and arginyl residues with respect to that employed for all the remaining ones. For all the other residues, we have determined the conformational energy maps by appropriatly superposing the *complete* series of sub-maps obtained with *all* the possible combinations of the preselected stereochemical rotational angles χ^j of the side chains (the superposition being done by taking the lowest energy

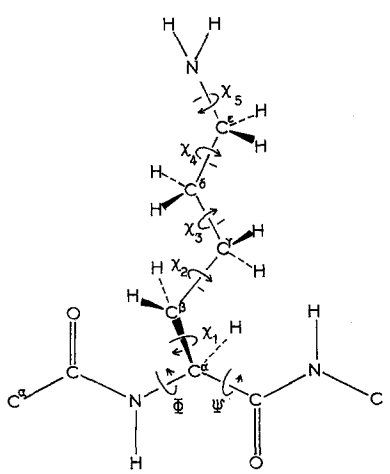


Fig. 1

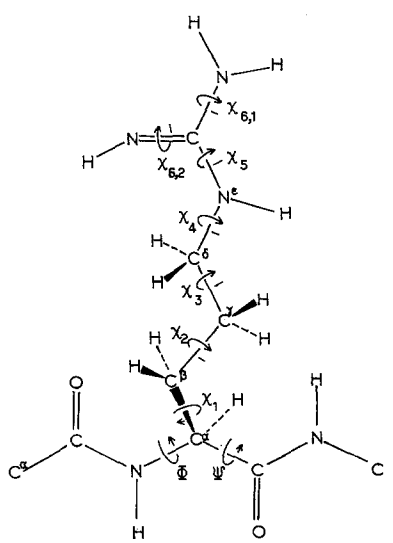


Fig. 2

Fig. 1. Standard conventions for studying the conformation of the lysyl residue [9]

Fig. 2. Standard conventions for studying the conformation of the arginyl residue [9]

points from each sub-map individually). Because of the very large number of such combinations for the here studied residues (81 possible combinations of χ^1, χ^2, χ^3 and χ^4 in the lysyl residue and 108 possible combinations of χ^1, χ^2, χ^3, χ^4 and χ^5 in the arginyl residue, with the three preselected values for each of them, not considering the possible rotations of the terminal amino group), this procedure was impracticable in the present case and a more empirical mode of approach had to be used. This consisted of selecting a *limited set* of the *most probable* such combinations from known experimental results on X-ray studies of globular proteins. A statistical evaluation of such combinations has already been undertaken by Chandrasekaran and Ramachandran [12] on the basis of the X-rays results for myoglobin [13], lysozyme [14, 15] and α-chymotrypsin [16] and it included 39 lysyl and 18 arginyl residues. Owing to the kindness of a number of colleagues we have been able to enlarge this statistics with (unpublished) data from ribonuclease-*S* (Dr. Wyckoff, Yale University) (see also [17]), carboxypeptidase-*A* (Drs. Lipscomb and Reeke, Harvard University) (see also [18, 19]), erythrocruorin (Dr. Huber, Max-Planck-Institute, Munich) (see also [20]), and oxyhaemoglobin (Dr. Perutz, Cambridge) (see also [21, 22]), so that it involved 81 sets of experimental values of the χ's for the lysyl residue and 38 such sets for the arginyl residue.

For the lysyl residue the three by far most common combinations of the χ's are:

$\chi^1 = 300°$	$\chi^2 = 180°$	$\chi^3 = 180°$	$\chi^4 = 180°$	Number: 29,
$\chi^1 = 180°$	$\chi^2 = 180°$	$\chi^3 = 180°$	$\chi^4 = 180°$	Number: 19,
$\chi^1 = 180°$	$\chi^2 = 60°$	$\chi^3 = 180°$	$\chi^4 = 180°$	Number: 6.

Because we know from previous studies that χ^1 is the most important parameter in the determination of the overall contour of the general conformational energy map of the residues, we have included in our computations one combination of the χ's involving $\chi^1 = 60°$. The remaining χ's in this combination are $\chi^2 = \chi^3 = \chi^4 = 180°$ and it occurs twice in the experimental results on the lysyl residue. By limiting ourselves to these four combinations we are taking into consideration 2/3 of all the observed ones.

For the arginyl residue, the experimental indications are somewhat less clear-cut, the selectivity of the combinations being less pronounced. The two most numerous ones are:

$\chi^1 = 300°$	$\chi^2 = 180°$	$\chi^3 = 180°$	$\chi^4 = 180°$	Number: 6,
$\chi^1 = 180°$	$\chi^2 = 180°$	$\chi^3 = 180°$	$\chi^4 = 180°$	Number: 5.

Other combinations of the χ's occur only once or twice. For reasons indicated above in connection with the lysyl residue, we have carried out calculations also for one set with $\chi^1 = 60°$ (and the remaining χ^1's $= 180°$; this combination occurs twice).

The guanidyl group itself, at the extremity of the side chain was considered to be planar. χ^5 was taken uniformly equal to zero, the other possibility, $\chi^5 = 180°$, being expected not to modify to any appreciable degree the aspect of the conformational map. This assumption of planarity of the guanidyl group is verified by Extended Hückel calculations on this group [23] and by X-ray results for small arginine compounds [24] and for the afore mentioned proteins (in which $\chi^5 = 0°$

is appreciably more frequent than $\chi^5 = 180°$). It may also be noted that the guanidyl group is, in these most popular combinations of the side chain rotational angles, coplanar with the two preceding bonds, C^γ—C^δ, C^δ—N^ε ($\chi^4 = 180°$).

As a general comment, it may be remarked that because of the sequence of a few —CH_2— groups in the side chain, the sequences of the χ's have frequently the repeating values of 180°. The situation corresponds to a staggered and extended conformation of the side-chain which should thus have a tendancy to point away from the backbone and to avoid hydrogen bonding with it, in difference to the case with residues having shorter polar side chains (for an elaboration of this point see Ref. [25]).

Results and Discussion

A. The Lysyl Residue

Figs. 3, 4, 5, and 6 represent the different conformational sub-maps for the lysyl residue as obtained with the four different combinations of the side chain rotational angles. On each sub-map the isoenergy lines are drawn with respect to the individual minimum of this sub-map. The global minimum is associated with the $\chi^1 = 60°$ sub-map, the individual minima of the three remaining sub-maps being 0.9 Kcal/mole, 1 Kcal/mole and 1,4 Kcal/mole above that of the global minimum for the $\chi^1 = 300°$ sub-map, $\chi^1 = 180°$, $\chi^2 = 180°$ sub-map and

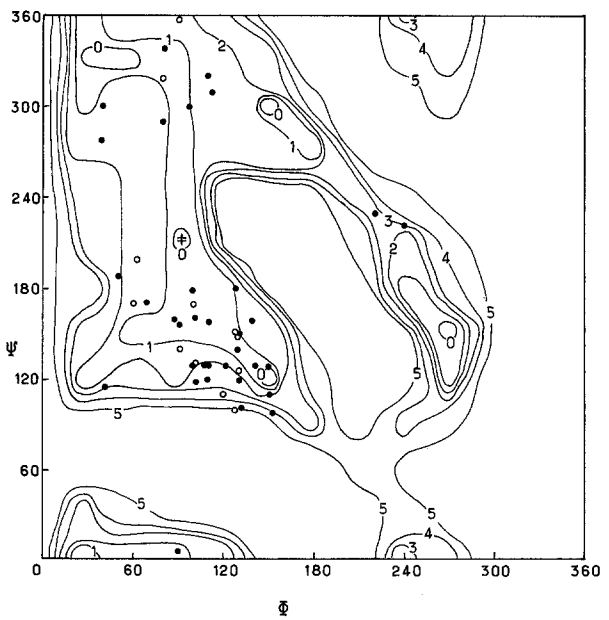


Fig. 3. Conformational energy sub-map for the lysyl residue corresponding to (χ^1, χ^2, χ^3, χ^4) = (300°, 180°, 180°, 180°). ∓ the minimum of the sub-map, ● conformations of the lysyl residues in globular proteins with this set of χ's, ○ conformations of other lysyl residues in globular proteins with $\chi^1 = 300°$

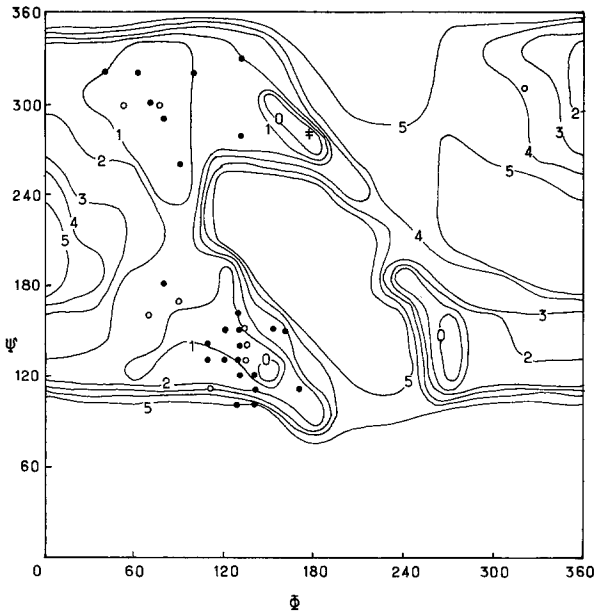


Fig. 4. Conformational energy sub-map for the lysyl residue corresponding to (χ^1, χ^2, χ^3, χ^4) = (180°, 180°, 180°, 180°). ∓ the minimum of the sub-map, ● conformations of the lysyl residues in globular proteins with this set of χ's, ○ conformations of other lysyl residues in globular proteins with $\chi^1 = 180°$

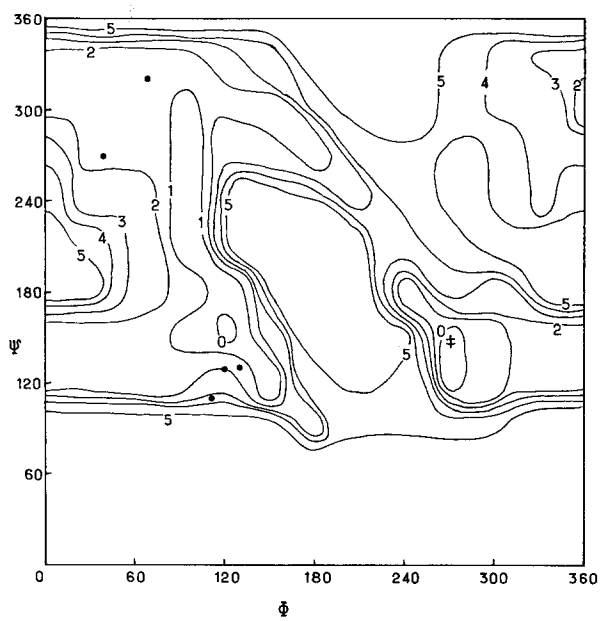


Fig. 5. Conformational energy sub-map for the lysyl residue corresponding to (χ^1, χ^2, χ^3, χ^4) = (180°, 60°, 180°, 180°). ∓ the minimum of the sub-map, ● conformations of the lysyl residues in globular proteins with this set of χ's, ○ conformations of other lysyl residues in globular proteins with $\chi^1 = 180°$

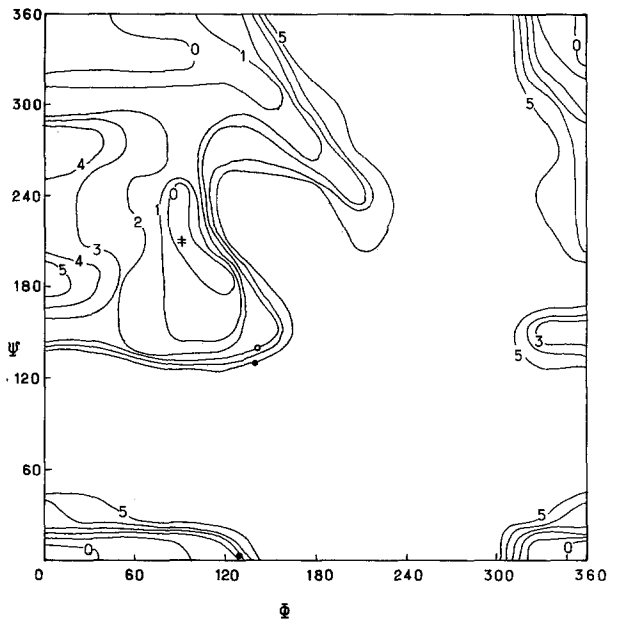


Fig. 6. Conformational energy sub-map for the lysyl residue corresponding to $(\chi^1, \chi^2, \chi^3, \chi^4) = (60°, 180°, 180°, 180°)$. ∔ the minimum of the sub-map, ● conformations of the lysyl residues in globular proteins with this set of χ's, ○ conformations of other lysyl residues in globular proteins with $\chi^1 = 60°$

$\chi^1 = 180°, \chi^2 = 60°$ sub-map, respectively. On each of the sub-maps the experimental conformations found in the above-quoted globular proteins corresponding to the very combination of the χ's utilized in the calculations are indicated by •, the remaining conformations with the same value of χ^1 but other values for the remaining χ's by *o*.

Fig. 7 represents the general conformational energy map for the lysyl residue, obtained from the superposition of these sub-maps by taking for each Φ, Ψ the lowest energy points, with respect to the global minimum, from the available sub-maps. The figure contains also the experimentally known conformation of the lysyl residue in globular proteins and includes beside those from the already quoted proteins, results from subtilisin (kindly communicated by Dr. Wright, Cambridge, England) (see also [26]), rubredoxin (kindly communicated by Dr. Jensen, University of Washington, Seattle) (see also [27]) and considerations on α-lactalbumin [28], proteins for which only the Φ and Ψ angles were available (or inferred).

The principal conclusions which may be drawn from these results are:

1) The comparison of the four sub-maps confirms that the general contour of the zone of conformational stability, within a given energy limit (5 Kcal/mole in our figures) is determined primarily by the value of χ^1. The comparison of the two sub-maps with $\chi^1 = 180°$ but with different χ^2's (180° or 60°) indicates, however, that the value of χ^2 influences the "fine structure" of this zone, in particular the contours of the low isoenergy curves.

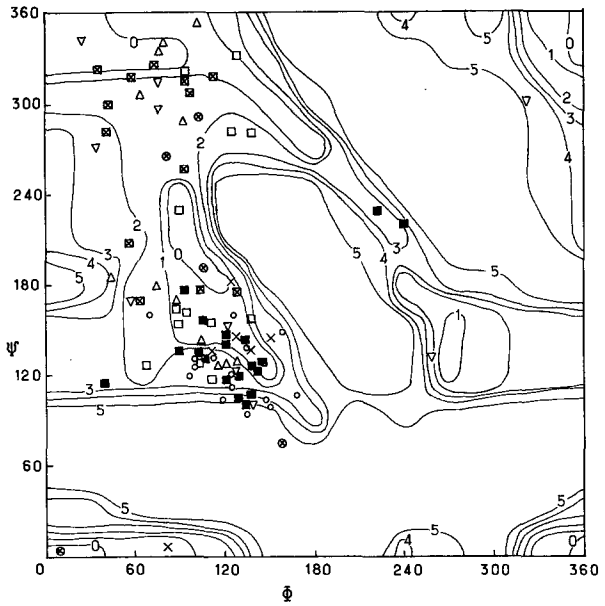


Fig. 7. Conformational energy map for the lysyl residue. Conformations of the residues in: ○ lysozyme, ● myoglobin, □ α-lactalbumin, ▽ ribonuclease-*S*, △ carboxypeptidase *A*, ⊠ α-chymotrypsin, ⊗ subtilisin, × erythrocruorin, ■ oxyhaemoglobin, ▲ rubredoxin

2) It is visible that in spite of their somewhat higher individual minima, the conformations corresponding to $\chi^1 = 300°$ and 180° have larger zones of low energy than the conformations corresponding to $\chi^1 = 60°$ and will thus have a larger *probability* of occurrence. They are in fact more populated.

3) The agreement between theory and experiment is extremely satisfactory, each group of experimental conformations lying well within the corresponding theoretical limits and in the vicinity of the local minima. Truly some of these local minima, in particular the one located in Figs. 3, 4, and 5 towards $\Phi, \Psi \approx 270°$, 140° are seldom populated. The significance of this situation will be considered in connection with the general conformational energy map.

4) This general conformational energy map merits a more detailed discussion. In the first place it is interesting to indicate the position of the energy minima. The global minimum occurs at $\Phi = 90°$, $\Psi = 210°$, the associated low energy zone extending till the conformation of Mizushima's seven-membered hydrogen ring ($\Phi, \Psi = 100°, 240°$), which we have previously designed by the symbol H′-7 [1, 3].

Several local minima occur at about 1 Kcal/mole above the global one. They are found towards:

a) $\Phi = 0°$, $\Psi = 0°$, corresponding to the fully extended form.

b) $\Phi = 270°$, $\Psi = 140°$, corresponding to the seven-membered hydrogen bonded conformation H-7 [1, 3].

c) $\Phi = 110° - 120°$, $\Psi = 120° - 130°$ corresponding to the R_α conformation.

The minima *a* and *c* are associated with large zones of very low energy. In particular the 0° Kcal/mole isoenergy contour of the minimum *a* extends into the β region.

The agreement between theory and experiment is again extremely satisfactory on this level too, practically all the experimental points falling within the theoretical stability contours and the great majority of them being concentrated within the low energy curves. A particular remark needs, however, to be made about the local minimum H−7. There seems to be only one experimental point corresponding to this conformation (with two other points situated in the neighbouring zone around $\Phi = \Psi = 240°$). Two reasons may explain this situation. In the first place, the relatively small area enclosed by, say, the 1 Kcal/mole curve of this minimum as compared with the similar limit of the other local minima indicates that the H−7 minimum must be associated with a relatively very small *probability* of occurrence. The situation is very similar to the one which we have considered recently in details in connection with the conformational energy map for the alanyl residue and in which it has been substantiated by explicit calculations of such probabilities [29]. In the second place it is obvious that such seven-membered hydrogen bonded forms, whether H′−7 and H−7, are essentially specific for the dipeptides studied and can hardly occur in proteins. Their occurrence in model dipeptides seems amply confirmed by recent experimentations [30–34].

B. The Arginyl Residue

The three conformational sub-maps for this residue corresponding to the three most common combinations of the rotational angles χ^1, χ^2, χ^3, χ^4 are indicated in Figs. 8, 9 and 10. The global minimum is again associated with the

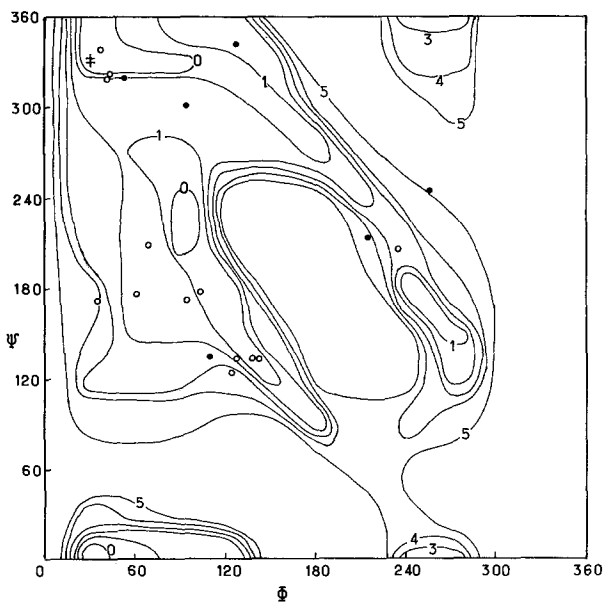


Fig. 8. Conformational energy sub-map for the arginyl residue corresponding to $(\chi^1, \chi^2, \chi^3, \chi^4, \chi^5) = (300°, 180°, 180°, 180°)$. ∓ the minimum of the sub-map, ● conformations of the arginyl residues in globular proteins with this set of χ's, ○ conformations of other arginyl residues in globular proteins with $\chi^1 = 300°$

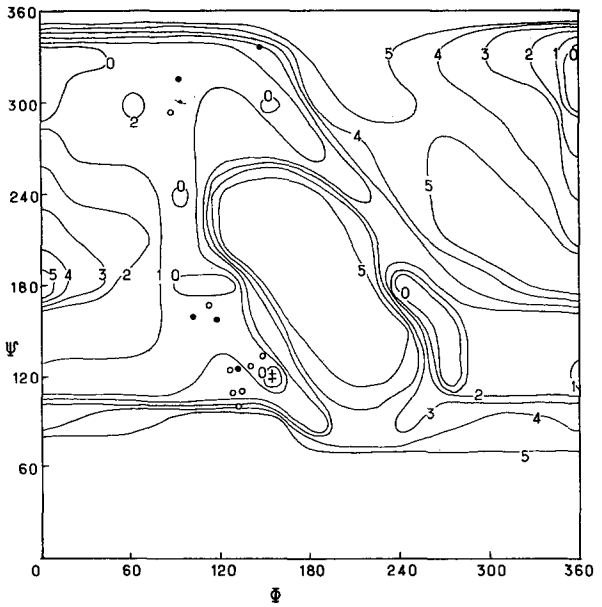


Fig. 9. Conformational energy sub-map for the arginyl residue corresponding to $(\chi^1, \chi^2, \chi^3, \chi^4, \chi^5) = (180°, 180°, 180°, 180°)$. ∓ the minimum of the sub-map, ● conformations of the arginyl residues in globular proteins with this set of χ's, ○ conformations of other arginyl residues in globular proteins with $\chi^1 = 180°$

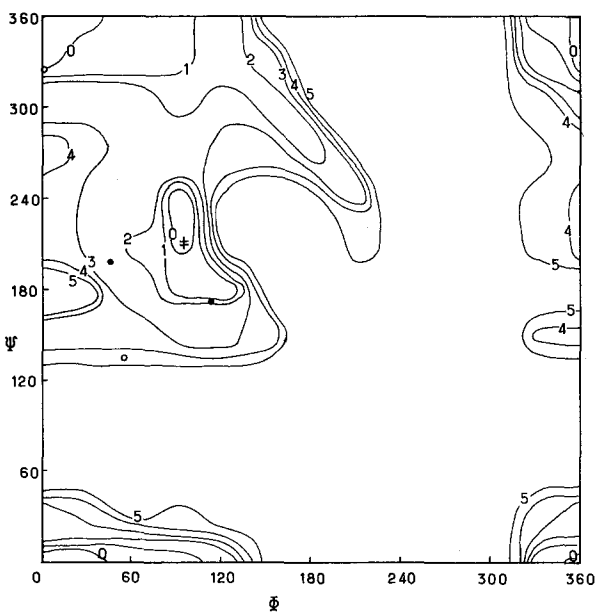


Fig. 10. Conformational energy sub-map for the arginyl residue corresponding to $(\chi^1, \chi^2, \chi^3, \chi^4, \chi^5) = (60°, 180°, 180°, 180°)$. ∓ the minimum of the sub-map, ● conformations of the arginyl residues in globular proteins with this set of χ's, ○ conformations of other arginyl residues in globular proteins with $\chi^1 = 60°$

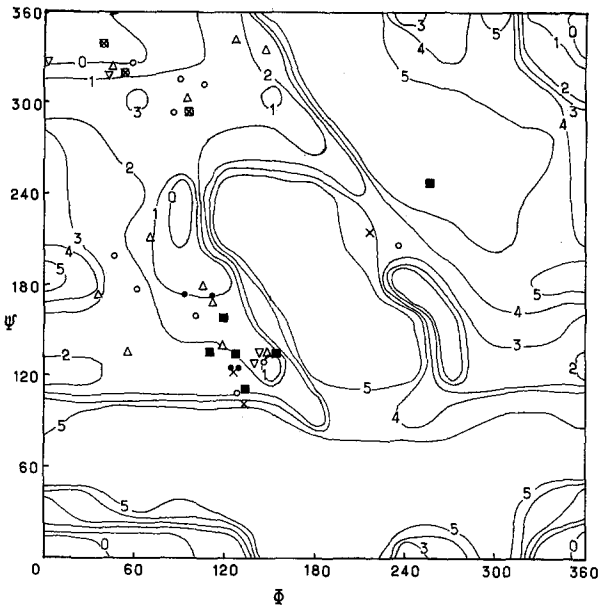


Fig. 11. Conformational energy map for the arginyl residue. Conformations of the residues in: ○ lysozyme, ● myoglobin, □ α-lactalbumin, ▽ ribonuclease-*S*, △ carboxypeptidase *A*, ⊠ α-chymotrypsin, × subtilisin, ⊗ erythrocruorin, ■ oxyhaemoglobin, ▲ rubredoxin

$\chi^1 = 60°$ sub-map, the individual minima of the two remaining sub-maps being 0.75 Kcal/mole and 1.3 Kcal/mole above that of this global minimum for the $\chi^1 = 300°$ and $\chi^1 = 180°$ sub-maps, respectively. Fig. 11 represents the general conformational energy map for that residue.

The principal conclusions which can be drawn from the examination of all these results are practically identical with those drawn from the study of the lysyl residue. It is again visible that in spite of their somewhat higher individual minima, the conformations corresponding to $\chi = 300°$ and 180° have *larger* zones of low energy than the conformations corresponding to $\chi^1 = 60°$ and will thus have a large probability of occurrence. They are again in fact more populated. As to the general conformational map, its great analogy (although not identity) with that of the lysyl residue is evident. The global minimum occurs again at $\Phi = 90°$, $\Psi = 210°$ and the associated low energy zone extends till the H′ − 7 conformation. A large zone of local minimum, located towards $\Phi = 0$–$60°$, $\Psi = 330$–$360°$ is again associated with the fully extended and the β-forms. There is again a local minimum in the R_α region and in the H − 7 region, the last one not being populated again, although two experimental conformations occur in its neighbourhood.

Conclusion

In conclusion it may be said that the procedure adopted here for the construction of the conformational energy map of the lysyl and arginyl residues, the largest among the amino-acid residues and which for practical reasons had therefore to

be treated less completely than the simpler ones, appears to lead to quite satisfactory results. In particular the stable conformational zones are correctly delimited. This result reconfirms the predominant significance of the rotational angle χ^1 in this respect, the role of the remaining χ's and in particular of χ^2 being nevertheless far from negligible as concerns the detailed structure of these stability zones. It appears again that the results although they indicate with particular precision the most stable conformations of the model dipeptide, procure at the same time satisfactory contours for the stable regions in globular proteins, a more precise comparison with the experimental values needing nevertheless the taking into consideration, besides the energy values, of the probabilities of occurrence of the preferred conformational forms.

References

1. Maigret, B., Pullman, B., Dreyfus, M.: J. theoret. Biology **26**, 321 (1970).
2. — — Perahia, D.: Biopolymers **10**, 107 (1971).
3. Pullman, B., Maigret, B., Perahia, D.: Theoret. chim. Acta (Berl.) **18**, 44 (1970).
4. Maigret, B., Perahia, D., Pullman, B.: J. theoret. Biology **29**, 275 (1970).
5. — — — Biopolymers, in press.
6. Perahia, D., Maigret, B., Pullman, B.: Theoret. chim. Acta (Berl.) **19**, 121 (1970).
7. Maigret, B., Pullman, B., Caillet, J.: Biochem. biophysic. Res. Commun. **40**, 808 (1970).
8. Pullman, B.: Aspects de la chimie quantique contemporaine, R. Daudel and Pullman, A. Eds., Proceedings of an International Symposium hold in Menton, July 1970, CNRS p. 261 (1971).
9. Edsall, Flory, P. J., Kendrew, J. C., Liquori, A. M., Nemethy, G., Ramachandran, G. N., Scheraga, H. A.: Biopolymers **4**, 121 (1966).
10. Diner, S., Malrieu, J. P., Jordan, F., Gilbert, M.: Theoret. chim. Acta (Berl.) **15**, 100 (1969), and the references indicated there.
11. Gurskaya, G. V.: Structury aminokistlot, Izdatel'stvo "Nauka", Moscow (1966).
12. Chandrasekaran, R., Ramachandran, G. N.: Reported in the Symposium on: Studies of conformational states of biopolymers, at the III. International Biophysics Congress, Cambridge, Mass. USA, September 1969.
13. Watson, H. C.: Progress in stereochemistry **4**, 299 (1969).
14. Phillips, D. C.: Proc. nat. Acad. Sci. USA **57**, 484 (1967).
15. Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C., Sarma, V. R.: Proc. Roy. Soc. (London) B **167**, 365 (1967).
16. Birktoft, J. J., Mathews, B. M., Blow, D. M.: Biochem. biophysic. Res. Commun. **36**, 131 (1969).
17. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B., Richards, F. M.: J. biol. Chemistry **245**, 305 (1970).
18. Reeke, G. N., Hartsuck, J. A., Ludwig, M. L., Quiocho, F. A., Steitz, T. A., Lipscomb, W. N.: Proc. nat. Acad. Sci. USA **58**, 2220 (1967).
19. Lipscomb, W. N., Hartsuck, J. A., Reeke, G. N., Quiocho, F. A., Bethge, P. H., Ludwig, M. L., Steitz, T. A., Muirhead, H., Coppola, J. C.: Brookhaven Symposia in Biology **21**, 24 (1968).
20. Huber, R., Epp, O., Formanek, H.: Naturwissenschaften **56**, 362 (1969).
21. Perutz, M. F., Muirhead, H., Cox, J. M., Goaman, L. C. G., Mathews, F. S., Gandy, E. L. M., Webb, L. E.: Nature **219**, 29 (1968).
22. — — — — Nature **219**, 131 (1968).
23. Ponnuswamy, P. K.: Thesis, University of Madras (1970).
24. Lakshminarayanan, A. V., Sasisekharan, V., Ramachandran, G. N.: Conformation of biopolymers, Ramachandran, G. N. ed., p. 61. New York: Interscience 1967.
25. Ramachandran, G. N., Chandrasekaran, R., Sarathy, K. P., Lakkaraju, R., in press. See also Ramachandran, G. N.: Acta crystallogr. B **25**, S 180 (1969).
26. Wright, C. S., Alden, R. A., Kraut, J.: Nature **221**, 235 (1969).
27. Herriott, J. R., Sieker, L. C., Jensen, L. H.: J. Mol. Biol. **50**, 391 (1970).

28. Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C., Hill, R. L.: J. molecular Biol. **42**, 65 (1969).
29. Maigret, B., Pullman, B., Perahia, D.: J. theoret. Biology, in press.
30. Bistrov, V. F., Portnova, S. L., Tsetlin, V. I., Ivanov, V. T., Ovchinnikov, Y. A.: Tetrahedron **25**, 493 (1969).
31. — — Balashova, T. A., Tsetlin, V. I., Ivanov, V. T., Kostetzky, P. V., Ovchinnikov, Y. A.: Tetrahedron Letters **59**, 5225 (1969).
32. Avignon, M., Huong, P. V., Lascombe, J., Marraud, M., Néel, J.: Biopolymers **8**, 69 (1969).
33. — — Biopolymers **9**, 427 (1970).
34. Maraud, M., Néel, J., Acignon, M., Huong, P. V.: J. Chim. physique **67**, 959 (1970).

Prof. B. Pullman
Institut de Biologie Physico-Chimique
13, rue P. et M. Curie
F-75 Paris 5è, France