Journal of Computer-Aided Molecular Design, 1 (1987) 211–217
ESCOM

J-CAMD 018

Energy minimization and molecular dynamics studies of Asn-102 elastase

Bogdan Lesyng* and Edgar F. Meyer Jr.**

Texas A & M University, Department of Biochemistry and Biophysics, College Station, TX 77843, U.S.A.

Received 17 June 1987
Revised 1 October 1987
Accepted 26 October 1987

Key words: Molecular dynamics; Energy minimization; Serine proteases; Enzyme mechanisms

SUMMARY

Four isomeric forms of the Asn-102 PPE (D102N mutant according to the emerging protocol, [Knowles, Science, 236 (1987) 1252–1258]) have been investigated using energy minimization (EM) and molecular dynamics (MD) techniques. MD simulation data for 175 ps are reported for each form (in total 700 ps for about 2500 atoms). The His-57 N_ε-protonated forms are calculated to be more stable than the N_δ-protonated ones. The active site region of the most stable form is very similar to that found in the D102N rat trypsin enzyme [Craik et al., Science, 237 (1987) 909–913]. Conformations of the active sites and their hydrogen bond patterns are presented for each of these forms and are compared with the structure of the native enzyme active site. The pH dependent activity of the D102N derivative is discussed.

INTRODUCTION

The serine proteases are hydrolytic enzymes characterized by the structurally homologous juxtaposition of a catalytic triad composed of residues: serine (Ser-195), histidine (His-57) and aspartate (Asp-102). While the general features of events leading up to substrate bond hydrolysis have been proposed for some time in terms of a 'charge relay' or 'proton relay' system, see, for example, Refs. 1–3, there are still disagreements regarding the role of the active site residues, and in particular of Asp-102, in the enzymatic activity of the enzymes [4].

By means of site specific mutagenesis, Craik et al. [5] created a mutant of rat trypsin with Asp-102 replaced by Asn (D102N form, according to the emerging protocol [6]). The mutant enzyme showed chemical reactivity of the histidine and serine residues as well as substrate binding, but the catalytic rate was reduced by a factor of 10^4 at physiological pH. At more alkaline conditions (pH

* On leave from the University of Warsaw, Department of Biophysics, Poland.
**To whom correspondence should be addressed.

$\approx$ 10) reactivity was restored to 15%. Crystal structures of rat trypsin at bracketing pH values confirmed the retention of structural homology [7]. Even at high resolution, however, X-ray crystallographic methods are hard pressed to distinguish between the N and O atoms in Asn-102, on the basis of electron density populations. A plausible assignment was made of atomic positions of the $>C=O$ atoms on the basis of an optimum hydrogen-bonding pattern.

There is a high degree of structural homology between trypsins and elastases, as well as ready access to hich resolution atomic coordinates of porcine pancreatic elastase (PPE) in native and a variety of complexed states [8–11]. Due to the availability of this extensive data base for PPE, we undertook the energy minimization (EM) and molecular dynamics (MD) simulation of asparagine-modified PPE.

The purpose of these calculations was:

(1) to analyse possible conformational forms of D102N PPE, and in particular to determine if the structure of the active site found in the crystalline state [7] is the most energetically favorable, so that the crystallographic results can be generalized for other D102N mutants of these proteolytic enzymes;

(2) to delineate possible hydrogen-bonding patterns in the active site of D102N PPE; and

(3) to attempt to explain changes in enzymatic activity as a function of pH; in turn, this could provide additional insight regarding details of the enzymatic reaction mechanism for serine proteases.

METHODS

The crystal structure of PPE complexed with a trifluoroacetyl-dipeptide-dimethylanilide inhibitor [12] and containing 45 crystallographically resolved water molecules [11, 13], has been used as a coordinate base to generate models of D102N PPE derivatives. Each of the Asp-102 carbonyl oxygen atoms was alternately replaced with a $-NH_2$ group resulting in the generation of two different initial conformational states of D102N PPE (denoted here as the A and B forms). In the productive, hydrogen-bonding configuration, the His-57 imidazole ring is approximately coplanar with Ser-195 O_γ and Asp-102 O_δ, thus, these discrete orientations were considered as starting points for minimization and dynamics calculations. Subsequently, the His-57 N_δ-protonated imidazole ring was replaced by its N_ε-protonated tautomer. These variations correspond to forms C and D, respectively.

The potential energies of these forms, as well as of the native enzyme have been minimized using the AMBER program system (version 2.0, [14]), employing the united atom potential model [15], a distance-dependent dielectric term, and a 12 Å atom-based cutoff distance. ESP electrostatic charges on the inhibitor atoms have been calculated using the SCF *ab initio* method [16], using the STO-3G basis set. The charges of the backbone atoms have been replaced with the standard ESP AMBER-library charges [14,15], which created a small nonbalanced charge (0.15 e), which has been distributed over the non-backbone atoms proportionally to their (ESP STO-3G) charges, as described in Ref. 17. This procedure satisfies in a qualitative way the electronegativity properties of the atoms (incremental charges are proportional to the initial charges). The chemical structure of the inhibitor with its partial charges is shown in Fig. 1.

A total of 100 steepest descent steps and 500–700 conjugate gradient cycles for each of the systems has been calculated during the minimization procedure. The minimization task was termi-

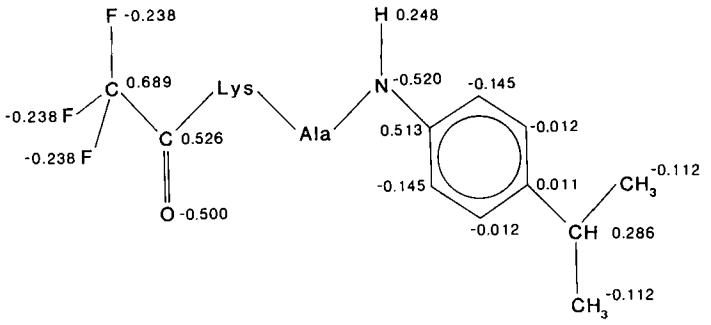

Fig. 1. Partial charges of the trifluoroacetyl-dipeptide-dimethylanilide inhibitor used in the EM and MD computations.

nated when the r.m.s. potential energy gradients were less than 0.3 kcal/mol Å. MD simulations for the D102N derivatives were performed after the minimization studies.

Prior to MD data collection, the model systems were thermalized and equilibrated using the following procedure:

Atoms were assigned initial velocities from a Maxwellian distribution at a temperature 285 K, i.e., below the reference, thermal bath temperature (300 K). The dynamics were simulated in the microcanonical ensemble 15 times for 0.2 ps (in total 3 ps for each of the structures), using a new set of random velocities each time. At this stage, additional 3.0 ps simulations in the microcanonical ensemble were calculated, which showed that the temperature of the systems was stabilized at 290–295 K.

After thermalization, the systems have been coupled to a thermal bath with a temperature relaxation time $\tau = 0.05$ ps. Equilibration was started using the restart files from the thermalization stage. This process was continued at a constant temperature of 300 K for an additional 50 ps. The potential energy of the systems stabilized after 25–30 ps. Longer equilibration periods were employed to reduce the energy error to as small a value as was possible during the data collection procedure. Before starting the data collection process the final equilibration coordinates were displayed on the PS-300 graphics system and superimposed with the corresponding crystallographic model. The molecules conserved their shape and no pronounced backbone deformations were noticed.

All bond lengths during the equilibration and data collection periods were constrained using the SHAKE algorithm [18], with a very small tolerance 0.0001 Å. Basic simulation calculations encompassing a period of 175 ps followed the thermalization and equilibration periods for all derivatives. The equations of motion were integrated with a time step of 0.001 ps (1.0 fs).

RESULTS AND DISCUSSION

The minimization procedure gave the following results: $U_C < U_A < U_D < U_B <$, with the potential energy difference of 0.0, 10.4, 14.8 and 19.2 kcal/mol, measured with respect to the C-form. The minimization itself does not provide overly reliable information on the stability of macromolecules, and in particular on the stability of proteins, because the conformational space of proteins exhibits a large number of thermally accessible minima; from this point of view proteins resemble glasses to some extent [19]. These minima are explored in the course of the MD simulations.

TABLE 1
THE MOLECULAR DYNAMICS ENERGY DATA (kcal/mol). <E> IS THE TOTAL ENERGY AVERAGED OVER THE 175 ps SIMULATION PERIOD, δE_{RMS} IS ITS RMS DEVIATION. Δ <E> IS THE ENERGY DIFFERENCE CALCULATED WITH RESPECT TO THE MOST STABLE FORM

Form	<E>	δE_{RMS}	Δ<E>
A	−2933.0	50.1	104.1
B	−2969.6	50.7	67.5
C	−3037.1	51.7	0.0 (Ref.)
D	−3032.8	50.7	4.3

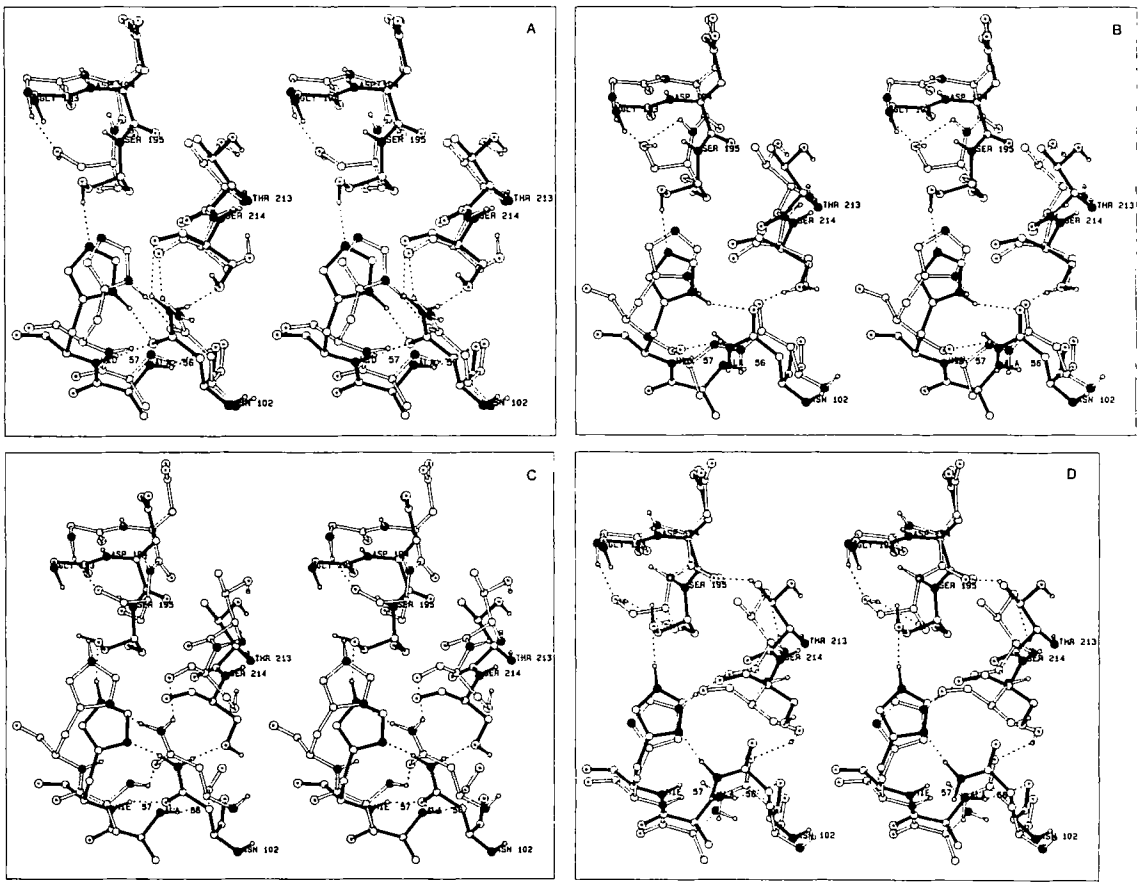

Fig. 2. Structure of the D102N PPE active site. The energy minimization results are shown as solid, black lines (bonds); the gradient norm is lower than 0.3 kcal/mol Å. The last 'snapshots' of the 175 ps simulation periods are shown with open, white bonds. Nitrogen atoms are represented as filled black circles. Oxygen atoms contain a 'dot' in the circle. A, B, C and D denote the A-, B-, C- and D-derivative forms (description in the text). HID-57 and HIE-57 denote the N_δ- and N_ϵ-protonated His-57 residue, respectively.

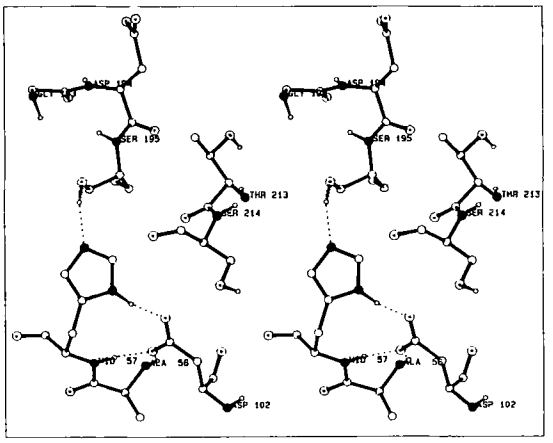

Fig. 3. Structure of the PPE (Asp-102) active site after energy minimization to a gradient norm less than 0.3 kcal/mol Å. Denotations as in Fig. 2.

The total energies of the A-, B-, C- and D-forms, averaged over the 175 ps MD trajectories, are given in Table 1. Figs. 2A, B, C and D show the structures of the active sites of these derivatives (minimization results vs. final MD snapshots, as 'before and after' representations). The conformational arrangement of the native enzyme active site is shown for comparison in Fig. 3.

According to the dynamic data, the His-57 N_ε-protonated C-derivative is the most stable one, and is therefore chosen to be the reference systems. The His-57 N_δ-protonated derivatives (A and B) are much less stable than the His-57 N_ε-protonated ones (C and D). Energies for the A- and B-derivatives are 67-104 kcal/mol higher than those corresponding to the C-derivative (this represents 1–2 times the r.m.s. deviation of the total energy (~ 50 kcal/mol), Table 1). The energy of the D-derivative is close to the energy of the C-derivative (~ 4 kcal/mol).

The energy dispersion is an important parameter when comparing the mean energy values. Therefore, we shall attempt to provide an independent estimate of the r.m.s. energy fluctuation based on experimental thermodynamical data.

Elastase is a globular protein with a molecular weight of about 30 000 Da. Heat capacities (C_p) of globular proteins are in the range of 0.3–0.35 cal g^{-1} deg^{-1} [20,21]. Using a general expression for the mean square fluctuation of the energy (see, for example, Ref. 22),

$$<\Delta E^2> = kT^2mC_v$$

with the following parameters: $T = 300$ K, $m = 5.04 \times 10^{-20}$ g/molecule, $C_p \approx C_v \approx 0.325$ cal g^{-1} deg^{-1}, we get $<\Delta E^2>^{1/2} \approx 42$ kcal/mol. This estimates the energy fluctuation per molecule expressed in kcal/mol. It is close to the MD δE_{RMS} values reported in Table 1. Because we can estimate the energy dispersion in our systems, we thus can describe the energy relationships in the comparative terms: 'close', 'different', 'greater' or 'smaller'.

Although the relatively large energy difference between the A- and B-, and C- and D-forms cannot be explained solely on the basis of structural differences in the active sites (the active sites interactions are responsible only for a fraction of the total energy), the interactions in the active sites reflect qualitatively the relative stabilities. More well-localized hydrogen bonds exist in the active sites of the C- and D-derivatives than in the sites of the A- and B-forms (see Fig. 2). The hy-

drogen-bonding pattern in the C-form, which stabilizes the position of the His-57 imidazole ring (minimization data), is identical to that found in the crystal state of D102N rat trypsin [7]. In general, the differences in the total energies of the systems are equal to the mean potential energy differences (the kinetic energy is the same with an accuracy of 0.1 kcal/mol). The relatively large energy difference between A- and B- vs. C- and D-forms is determined mainly by the long range electrostatic contribution; approximately 70% of the energy difference can be found in the electrostatic term. This suggests that the unfavorable perturbations in the active sites of the A- and B-forms propagate over the rest of the molecule, inducing overall destabilization effects.

An interesting MD event is the change of the hydrogen-bonding pattern in the active site of the D-derivative during the MD simulation. The His-57 imidazole ring flipped over* about the C_β-$C\gamma$ bond. In addition, the positions of Thr-213 and Ser-214 shifted significantly relative to their initial positions. These conformational changes resulted in the breaking of all the initially formed hydrogen bonds and in the creation of a larger number of new hydrogen bond interactions. The most significant change appears to be the disruption of a bond between HN_ε of His-57 and O_γ of Ser-195, and the formation of a new hydrogen bond between HN_ε and the oxygen atom of the Thr-213 carbonyl, showing hydrogen-bonding interactions not usually found in 'native' serine proteases.

Whereas the conformation of the active site in D102N trypsin is very similar to the structure of the PPE C-derivative, the conformational arrangement of the active site in D and the stability of this form was not suggested by the crystal structure data. In addition, the significantly higher energy of the His-57 N_δ-protonated forms (A, B) is a result which could not have been easily predicted without employing MD simulations. Our results may therefore be taken as an extension of the available crystallographic data [7].

An interesting result of these dynamical calculations is the observation that the hydroxyl group of Ser-195 can itself fit into the 'oxyanion hole' (region between the N-H groups of residues 193 and 195) (see Fig. 2), even though this relationship has thus far not been reported in available crystal structures of native proteases.

Regarding the enzymatic activity of D102N PPE, one can easily see that the most stable C- and D-derivatives cannot be biologically active at neutral pH because the N_ε-protonated His-57 residue is unable to participate in proton extraction from Ser-195. Note that native proteases are N_δ-protonated (see Refs. 1–3, 23, cf. Fig. 3). The biological activity of D102N PPE may be restored at higher pH when the HN_ε proton of His-57 dissociates. The geometry of the mutant active site in this case may be similar to that of the native enzyme. At these higher pH values, it is possible for the Asn-102 NH_2 group to participate in a hydrogen-bonding interaction with N_δ of His-57, fixing the position of the imidazole ring. The His-57 N_ε nitrogen atom can provide its lone electron pair as the potential acceptor, mediating the extraction of the Ser-195 hydroxyl proton in the crucial enzymatic step leading to nucleophilic attack on the substrate.

Due to the close similarity of the active sites in PPE, rat trypsin, and other serine proteases, we believe that our results are generally applicable for the homologous familiy of serine proteases.

*Note added in proof: We extended the simulation period for the D-form to 225 ps. Also, by means of program MDKINO (S. Swanson, Texas A&M Univ, unpublished), it was possible to view consecutive 0.1 ps snapshots of the active site region. The structure of the region of the D-form was relatively stable during first 100 ps of the simulation. However, during the period of 100-225 ps of the simulation we observed four 180° flips of the His-57 imidazole ring (2 forwards and 2 backwards flips, each of which takes about 1-1.5 ps), and one 360° flip completed in about 2.5 ps.

Our theoretical results could be verified by properly planned experiments (neutron diffraction, NMR spectroscopy and additional X-ray structure analysis). In particular it should be determined if the D-derivative exists under physiological conditions.

Our simulations can be extended using the thermodynamic-cycle-perturbational method (see, for example, Refs. 24,25), to calculate free energy differences between the D102N forms. With the knowledge of the most stable conformations, such calculations can be more easily initiated.

ACKNOWLEDGEMENTS

We thank Drs. Gail Carlson and Stanley Swanson for computational collaborations, we thank Dr. Charles Craik for providing results prior to publication. Financial support has been provided by the National Science Foundation, the Office of Naval Research and the Robert A. Welch Foundation. Computational resources were provided by Vice Provost John Dinkel, HARC and NEC. B.L. wishes to acknowledge travel support from the Ministry of Science and Higher Education (Poland) within the project CPBP 01.06.

REFERENCES

1 Allen, L.C., In Weinstein, H. and Green, J.P. (Eds.) Quantum Chemistry in Biomedical Sciences, Annals of the New York Academy of Sciences, Vol. 367, New York, 1981.
2 Shibata, M., Kieber-Emmons, T., Dutta, S.and Rein, R., Int. J. Quantum Chem. QBS, 8 (1981) 409–417.
3 Bizzozero, S.A. and Dutler, H., Bioorg. Chem., 10 (1981) 46–62.
4 Warshel, A. and Russell, S., J. Am. Chem. Soc., 108 (1986) 6569–6579.
5 Craik, C.S., Roczniak, S., Largman, C. and Rutter, J., Science, 237 (1987) 909–913.
6 Knowles, J.R., Science, 236 (1987) 1252–1258.
7 Sprang, S., Standing, T., Fletterick, R.J., Stroud, R.M., Finer-Moore, J., Xuong, N.-H., Hamlin, R., Rutter, W.J. and Craik, C.S., Science, 237 (1987) 905–908.
8 Meyer Jr., E.F. and Presta, L., In Burgen, A.S.V., Roberts, G.C.K. and Tute, M.S. (Eds.) Molecular Graphics and Drug Design, Elsevier Science Publishers, Amsterdam, 1986, pp. 307–321.
9 Meyer Jr., E.F., Presta, L. and Radhakrishnan, R., J. Am. Chem. Soc., 107 (1985) 4091–4093.
10 Meyer Jr., E.F., Radhakrishnan, R., Cole, G.M. and Presta, L., J. Mol. Biol., 189 (1986) 533–539.
11 Hughes, D.L., Sieker, L.C., Bieth, J., Dimicoli, J.-L., J. Mol. Biol., 162 (1982) 645–655.
12 With an eye to pursuing MD perturbation calculations, the original CF_3 terminus at the anilide ring was here substituted with -$CH(CH_3)_2$.
13 Protein Data Bank, See Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Schimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535–542.
14 Singh, U.C., Weiner, P.K., Caldwell, J. and Kollman, P., AMBER: Assisted Model Building with Energy Refinement Package, Version 2, Department of Pharmaceutical Chemistry, University of California, San Francisco, 1985.
15 Weiner, S.J., Kollman, P.A., Nguen, D.T. and Case, D.A., J.Comput. Chem., 7(1986) 230–252.
16 Singh, U.C. and Kollman, P.A., J. Comput. Chem., 5 (1984) 129–145.
17 Carlson, G., Molecular Dynamics Studies of Porcine Pancreatic Elastase: The Native State and Complexed to Two Trifluoroacetyl Dipeptide Anilide Inhibitors, Ph. D. Dissertation, Texas A & M University, 1987.
18 Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C., J. Comput. Phys., 23 (1977) 327–341.
19 Elber, R. and Karplus, M., Science, 235 (1987) 318–321.
20 Privalov, P.L., FEBS Lett. Suppl., 40 (1974) 140–153.
21 Suurkuusk, J., Acta Chem. Scand., B28 (1974) 409–417.
22 Callen, H.B., Thermodynamics, John Wiley and Sons, New York, 1960.
23 Bachovchin, W.W., Biochemistry, 25 (1986) 7751-7759.
24 Lybrand, T.P., McCammon, J.A. and Wipff, G., Proc. Nat. Acad. Sci. U.S.A., 83 (1986) 833–835.
25 Bash, P.A., Singh, U.C., Langridge, R. and Kollman, P.A., Science, 236 (1987) 564–568.