Theoretical Studies on the Bioactive Conformation of Nerve Growth Factor Using VBMC−A Novel Variable Basis Monte Carlo Simulated Annealing Algorithm for Peptides

Igor L. Shamovsky,[†,‡] Gregory M. Ross,[‡] Richard J. Riopelle,[‡] and Donald F. Weaver*[,†,‡]

Contribution from the Departments of Chemistry and Medicine, Queen's University, Kingston, Ontario, Canada K7L 3N6

Received April 4, 1996[⊗]

Abstract: Nerve Growth Factor (NGF) is an important neurotrophic protein implicated in Alzheimer's disease, epilepsy, and pain. Although the amino and carboxyl termini enable an NGF's recognition by its TrkA receptor, their conformations have not been resolved in recent crystallographic studies. Variable Basis Monte Carlo simulated annealing calculations are utilized to study low-energy conformational space of the NGF termini of three highly active and three inactive NGF analogues in order to determine their bioactive conformation. The complex of the NGF termini splits into two distinct moieties: a rigid region (residues 9-11 and 112′-118′) and a flexible loop (residues 1-8). The geometry of the rigid region, which is maintained by electrostatic interaction between Glu^{11} and $Arg^{118'}$, is conserved in active molecules only. The separation of the flexible loop from the rigid region is necessary in order to eliminate an influence of the loop on the biologically active conformation of the rigid region. Experimentally observed structure−activity relationships can be explained using this structural model.

1. Introduction

The neurotrophins are a family of structurally and functionally related proteins, including Nerve Growth Factor (NGF), Brain-derived Neurotrophic Factor (BDNF), Neurotrophin-3 (NT-3), and Neurotrophin-4/5 (NT-4/5). These proteins promote the survival and differentiation of diverse neuronal populations in both the peripheral and central nervous systems.[1] Neurotrophins exert these biological effects through specific interactions with various tyrosine kinase receptors (TrkA, TrkB, and TrkC):[2−9] TrkA only binds NGF;[2,3] TrkB binds BDNF, NT-3, and NT-4/5;[4−10] TrkC exclusively binds NT-3.[8,11] Since only cells expressing Trk-type receptors show functional responses upon neurotrophin binding,[12−20] the role of the coexpressed common neurotrophin receptor, termed $p75^{NTR}$, remains unclear.[15,16,18−24]

NGF, a 118 amino acid protein, is an extremely important neurotrophin, being implicated in the pathogenesis of Alzheimer's disease, epilepsy, and pain.[25] The binding of NGF to its receptors is determined by distinct sequences within its primary amino acid structure. The hairpin loop at residues 29−35 is responsible for recognition by $p75^{NTR}$,[16,26] while the amino and carboxyl termini are important binding determinants for

† Department of Chemistry.
‡ Department of Medicine.
⊗ Abstract published in *Advance ACS Abstracts,* September 1, 1996.

(1) (a) Hefti, F. H. *J. Neurosci.* **1986**, *6*, 2155−2162. (b) Hefti, F. H.; Weiner, W. J. *Annals Neurology* **1986**, *20*, 275−281. (c) Levi-Montalcini, R. *EMBO J.* **1987**, *6*, 1145−1154. (d) Barde, Y.-A. *Neuron* **1989**, *2*, 1525−1534. (e) Leibrock, J.; Lottspeich, A. H.; Hohn, A.; Hofer, M.; Hengerer, B.; Masiakowski, P.; Thoenen, H.; Barde, Y.-A. *Nature* **1989**, *341*, 149−152. (f) Maisonpierre, P. C.; Belluscio, L. S. S.; Squinto, S.; Ip, N. Y.; Furth, M. E.; Lindsay, R. M.; Yancopoulos, G. D. *Science* **1990**, *247*, 1446−1451. (g) Rosenthal, A.; Goeddel, D. V.; Nguyen, T.; Lewis, M.; Shih, A.; Laramee, G. R.; Nikolics, K.; Winslow, J. W. *Neuron* **1990**, *4*, 767−773. (h) Hohn, A.; Leibrock, J.; Bailey, K.; Barde, Y.-A. *Nature* **1990**, *344*, 339−341. (i) Götz, R.; Köster, R.; Winkler, C.; Raulf, F.; Lottspeich, F.; Schard, M.; Thoenen, H. *Nature* **1994**, *372*, 266−269. (j) Maness, L. M.; Kastin, A. J.; Weber, J. T.; Banks, W. A.; Beckman, B. S.; Zadina, J. E. *Neurosci. Biobehav. Rev.* **1994**, *18*, 143−159.

(2) Kaplan, D. R.; Hempstead, B. L.; Martin-Zanca, D.; Chao, M. V.; Parada, L. F. *Science* **1991**, *252*, 554−558.

(3) Klein, R.; Jing, S.; Nanduri, V.; O'Rourke, E.; Barbacid, M. *Cell* **1991**, *65*, 189−197.

(4) Soppet, D.; Escandon, E.; Maragos, J.; Middlemas, D. S.; Reid, S. W.; Blair, J.; Burton, L. E.; Stanton, B.; Kaplan, D. R.; Hunter, T.; Nikolics, K.; Parada, L. F. *Cell* **1991**, *65*, 895−903.

(5) Squinto, S. P.; Stitt, T. N.; Aldrich, T. H.; Davis, S.; Bianco, S. M.; Radziejewski, C.; Glass, D. J.; Masiakowski, P.; Furth, M. E.; Valenzuela, D. M.; DiStefano, P. S.; Yancopoulos, G. D. *Cell* **1991**, *65*, 885−893.

(6) Berkemeier, L. R.; Winslow, J. W.; Kaplan, D. R.; Nikolics, K.; Goeddel, D.; Rosenthal, A. *Neuron* **1991**, *7*, 857−866.

(7) Escandon, E.; Burton, L. E.; Szönyi, E.; Nikolics, K. *J. Neurosci. Res.* **1993**, *34*, 601−613.

(8) Lamballe, F.; Klein, R.; Barbacid, M. *Cell* **1991**, *66*, 967−970.

(9) Klein, R.; Lamballe, F.; Bryant, S.; Barbacid, M. *Neuron* **1992**, *8*, 947−956.

(10) Barbacid, M. *Oncogene* **1993**, *8*, 2033−2042.

(11) Vale, R. D.; Shooter, E. M. *Methods Enzymol.* **1985**, *109*, 21−39.

(12) Glass, D.; Nye, S.; Hantzopoulos, P.; Macchi, M.; Squinto, S.; Goldfarb, M.; Yancopoulos, G. *Cell* **1991**, *66*, 405−413.

(13) Weskamp, G.; Reichardt, L. F. *Neuron* **1991**, *6*, 649−663.

(14) Ibáñez, C. F.; Ebendal, T.; Rersson, H. *EMBO J.* **1991**, *10*, 2105−2110.

(15) Drinkwater, C. C.; Suter, U.; Angst, C.; Shooter, E. M. *Proc. Royal Soc. London B Biol. Sci.* **1991**, *246*, 307−313.

(16) Ibáñez, C. F.; Ebendal, T.; Barbany, G.; Murray-Rust, J.; Blundell, T. L.; Persson, H. *Cell* **1992**, *69*, 329−341.

(17) Jing, S.; Tapley, P.; Barbacid, M. *Neuron* **1992**, *9*, 1067−1079.

(18) Barker, P. A.; Shooter, E. M. *Neuron* **1994**, *13*, 203−215.

(19) Loeb, D. M.; Maragos, J.; Martin-Zanca, D.; Chao, M. V.; Parada, L. F.; Greene, L. A. *Cell* **1991**, *66*, 961−966.

(20) Rovelli, G.; Heller, R. A.; Canossa, M.; Shooter, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8717−8721.

(21) Bothwell, M. A. *Cell* **1991**, *65*, 915−918.

(22) Chao, M. V. *Neuron* **1992**, *9*, 583−593.

(23) Shih, A.; Laramee, G. R.; Schmelzer, C. H.; Burton, L. E.; Winslow, J. W. *J. Biol. Chem.* **1994**, *269*, 27679−27686.

(24) Windisch, J. M.; Marksteiner, R.; Schneider, R. *J. Biol. Chem.* **1995**, *270*, 28133−28138.

(25) (a) Ben Ari, Y.; Represa, A. *TINS* **1990**, *13*, 312−318. (b) McKee, A. C.; Kosik, K. S.; Kowal, N. W. *Ann. Neurol.* **1991**, *30*, 156. (c) Leven, G. R.; Mendel, L. M. *TINS* **1993**, *16*, 353−359. (d) Woolf, C. J.; Doubell, T. A. *Current Opinions Neurobiol.* **1994**, *4*, 525−534. (e) Rashid, K.; van der Zee, C. E. E. M.; Ross, G. M.; Chapman, C. A.; Stanisz, J.; Riopelle, R. J.; Racine, R. J.; Fahnestock, M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9495−9499. (f) McMahon, S. B.; Bennett, D. L. H.; Priestley, J. V.; Shelton, D. L. *Nature Medicine* **1995**, *1*, 774−780. (g) Cosgaya, J. M.; Latasa, M. J.; Pascual, A. *J. Neurochem.* **1996**, *67*, 98−104.

(26) Radziejewski, C.; Robinson, R. C.; DiStefano, P. S.; Taylor, J. W. *Biochemistry* **1992**, *31*, 4431−4436.

S0002-7863(96)01119-5 CCC: $12.00

recognition by the TrkA receptor.[23,27−35] Truncation of either the amino or carboxyl terminus of NGF produces less active NGF analogues; similarly most deletion or point mutations of the amino terminus also lead to NGF analogues with diminished activity.[23,29,30,32−35] On the other hand, the NGFΔ2-8 (NGF with residues 2-8 removed) and NGFΔ3-9 deletion mutants are almost as active as wild type NGF.[32] These NGF structure−activity relationships in combination with the considerable species variability of the amino acid sequence of the NGF termini (mouse, human, guinea pig, and snake)[36] are of potential value in understanding the NGF-TrkA interaction.

NGF exerts its biological activity as a noncovalent dimer.[33,35−38] Two 118 residue NGF monomers are dimerized by hydrophobic and van der Waals interactions between their three anti-parallel pairs of β-strands; consequently, the amino terminus of one NGF monomer and the carboxyl terminus of the other are spatially juxtaposed.[36] Furthermore, although a dimer has two pairs of termini, only one pair of termini is required for TrkA receptor recognition.[33,35] Accordingly, solving the conformation of the complex formed by the amino terminus of one monomer and the carboxyl terminus of the other in dimeric NGF is of fundamental relevance to understanding the interaction of NGF with TrkA.

The X-ray crystallographic three-dimensional structure of a dimeric mouse NGF (mNGF) has been reported recently.[36] However, within this structure, the amino terminus (residues 1-11) and the carboxyl terminus (residues 112-118) remain unresolved for both pairs of termini. Since the amino and carboxyl termini are crucial to NGF bioactivity as mediated via TrkA and because of the significance of NGF in multiple neurologic disease processes, the determination of the biologically active conformation of these termini is an important and challenging problem for computational chemistry.

The aim of this computational study was to ascertain the biologically active conformation of the TrkA receptor binding determinant of NGF. A comprehensive review of the literature identified an initial study set of 17 proteins which have been assessed for TrkA mediated activity using comparable experimental methods. From these 17 analogues, a subset of six (three active and three inactive) which reflects the range of structural diversity was selected for preliminary study. Accordingly, in the present study the low-energy conformations of the amino and carboxyl termini of six NGF analogues were obtained by means of Monte Carlo simulated annealing calculations. Three of these analogues retain full TrkA mediated activity: human NGF (hNGF),[23,29] mouse NGF (mNGF),[29,32] and mNGFΔ3-9 (deletion 3-9 of mNGF).[32] The other three analogues are biologically inactive: hNGF-H4D (a point mutation of hNGF in which His [H] in position 4 is replaced by Asp [D]),[23] mNGFΔ1-8,[29,32,34] and BDNF.[28] The results and conclusions obtained from these six analogues were then validated by being applied to the remaining 11 analogues of the initial study set.

The structural optimization of the amino and carboxyl termini of the NGF analogues under study required a rigorous search of the conformational space. As with other theoretical peptide studies, such calculations are plagued by a multiple-minima problem[39] which traditionally is addressed using either molecular dynamics, Monte Carlo approaches, or other algorithms. The Monte Carlo method, although exceedingly powerful, is limited by the excessive CPU time requirements on existing computers. To remedy this problem, we have devised a novel Variable Basis Monte Carlo simulated annealing algorithm for peptides (VBMC).

2. Methods.

2.1. Background to the VBMC Algorithm. The potential energy surface of a peptide has an enormous number of local minima,[39−41] but usually only the global energy minimum structure is required. Despite efforts to elaborate an efficient method of global energy minimization,[39−59] the multiple-minima problem is far from being solved. Although simulated annealing[42,45] guarantees convergence to the global minimum in principle, a reliable "cooling" schedule is far too slow for practical purposes, and faster "cooling" has to be used.[60] Within the simulated annealing approach, thermal molecular motion can be simulated either by Monte Carlo[42,49,61−63] or molecular

(27) Moore, J. B.; Shooter, E. M. *Neurobiology* **1975**, *5*, 369−381.
(28) Suter, U.; Angst, C.; Tien, C.-L.; Drinkwater, C. C.; Lindsay, R. M.; Shooter, E. M. *J. Neurosci*. **1992**, *12*, 306−318.
(29) Burton, L. E.; Schmelzer, C. H.; Szönyi, E.; Yedinak, C.; Gorrell, A. *J. Neurochemistry* **1992**, *59*, 1937−1945.
(30) Kahle, P.; Burton, L. E.; Schmelzer, C. H.; Hertel, C. *J. Biol. Chem.* **1992**, *267*, 22707−22710.
(31) Luo, Y.; Neet, K. E. *J. Biol. Chem*. **1992**, *267*, 12275−12283.
(32) Drinkwater, C. C.; Barker, P. A.; Suter, U.; Shooter, E. M. *J. Biol. Chem*. **1993**, *268*, 23202−23207.
(33) Treanor, J. J. S.; Schmelzer, C.; Knusel, B.; Winslow, J. W.; Shelton, D. L.; Hefti, F.; Nikolics, K.; Burton, L. E. *J. Biol. Chem*. **1995**, *270*, 23104−23110.
(34) Taylor, R. A. V.; Kerrigan, J. F.; Longo, F. M.; deBoisblanc, M.; Mobley, W. C. *Soc. Neurosci. Abs*. **1991**, *17*, 712.
(35) Burton, G.; Schmelzer, C.; Sadic, M.; Hefti, F.; Treanor, J. *Soc. Neurosci. Abs*. **1995**, *21*, 1061.
(36) McDonald, N. Q.; Lapatto, R.; Murray-Rust, J.; Gunning, J.; Wlodawer, A.; Blundell, T. L. *Nature* **1991**, *354*, 411−414.
(37) Ibäñez, C. F.; Ilag, L.; Murray-Rust, J.; Persson, H. *EMBO J*. **1993**, *12*, 2281−2293.
(38) Bothwell, M. A.; Shooter, E. M. *J. Biol. Chem*. **1977**, *23*, 8532−8536.
(39) Gibson, K. D.; Scheraga, H. A. In *Structure and Expression: Vol. 1. From Proteins to Ribosomes*; Sarma, M. H., Sarma, R. H., Eds.; Adenine Press: Guilderland, NY, 1988; pp 67−94.
(40) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem*. **1989**, *93*, 3339−3346.
(41) Piela, L.; Olszewski, K. A.; Pillardy, J. *J. Mol. Struct. (Theochem)* **1994**, *308*, 229−239.
(42) Kirkpatrick, S.; Gelatt Jr., C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(43) Scheraga, H. A. *Prog. Clin. Biol. Res*. **1989**, *289*, 3−18.
(44) Scheraga, H. A. *Pol. J. Chem*. **1994**, *68*, 889−891.
(45) Vanderbilt, D.; Louie, S. G. *J. Comput. Phys*. **1984**, *56*, 259−271.
(46) Simon, I.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 797−804.
(47) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc*. **1989**, *111*, 4379−4386.
(48) Li, Z.; Scheraga, H. A. *J. Mol. Struct. (Theochem)* **1988**, *179*, 333−352.
(49) Nayeem, A.; Vila, J.; Scheraga, H. A. *J. Comput. Chem*. **1991**, *12*, 594−605.
(50) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1391−1436.
(51) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283−1303.
(52) Purisima, E. O.; Scheraga, H. A. *J. Mol. Biol*. **1987**, *196*, 697−709.
(53) Somorjai, R. L. *J. Phys. Chem*. **1991**, *95*, 4141−4146.
(54) Zakharov, V. V. *Eng. Cybern. (Engl. Transl.)* **1970**, *4*, 637−642.
(55) Olszewski, K. A.; Piela, L.; Scheraga, H. A. *J. Phys. Chem*. **1992**, *96*, 4672−4676.
(56) Amara, P.; Hsu, D.; Straub, J. E. *J. Phys. Chem*. **1993**, *97*, 6715−6721.
(57) Gordon, H. L.; Somorjai, R. L. *J. Phys. Chem*. **1992**, *96*, 7116−7121.
(58) Raj, N.; Morley, S. D.; Jackson, D. E. *J. Mol. Struct. (Theochem)* **1994**, *308*, 175−190.
(59) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M. *J. Mol. Struct. (Theochem)* **1994**, *308*, 191−206.
(60) (a) Gidas, B. *J. Stat. Phys*. **1985**, *39*, 73−131. (b) Kushner, H. J. *SIAM J. Appl. Math*. **1987**, *47*, 169−185.
(61) (a) Wilson, S. R.; Cui, W.; Moscowitz, J. W.; Schmidt, K. E. *Tetrahedron Lett*. **1988**, *29*, 4373−4376. (b) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett*. **1988**, *229*, 317−324. (c) Brünger, A. T. *J. Mol. Biol*. **1988**, *203*, 803−816. (d) Kawai, H.; Kikuchi, T.; Okamoto, Y. *Protein Engineering* **1989**, *3*, 85−94. (e) Wilson, C.; Doniach, S. *Proteins* **1989**, *6*, 193−209.
(62) Snow, M. E. *J. Comput. Chem*. **1992**, *13*, 579−584.
(63) Shamovsky, I. L.; Yarovskaya, I. Yu.; Khrapova, N. G.; Burlakova, E. B. *J. Mol. Struct. (Theochem)* **1992**, *253*, 149−159.

dynamics[64,65] methods. The major problem of the Monte Carlo technique is that motion of a molecular system along a Markov chain is very slow.[41,43,49] The Monte Carlo technique, however, has advantages over molecular dynamics: (i) it does not require calculating forces; (ii) potential energy functions do not have to be continuous; (iii) much larger steps in conformational space can be utilized;[48,66] (iv) the additional iterative self-consistent SHAKE routine,[67] used in molecular dynamics simulations,[68] is not required. This has led to efforts to accelerate the Monte Carlo approach.[58,62,66,69,70] The main reason for inefficiency with the conventional Monte Carlo method[71] is a slow and cumbersome sequential change of configurational variables. To overcome these problems, lessons can be learned from molecular dynamics. In molecular dynamics[72] configurational variables move simultaneously, and the soft modes (defined as low-energy valleys on the potential energy surface) of the system explicitly direct these motions. Although there have been attempts to direct Monte Carlo configurational motion along the soft modes,[58,69,70] these improvements require calculating forces. An algorithm which would retain the advantages of the Monte Carlo method and direct configurational changes along the soft modes is thus required.[73]

The classical Monte Carlo−Metropolis method[71] consists of modeling a Markov chain containing subsequent molecular conformations. Each vector of independent structural parameters $\mathbf{x}_{i+1}$ is derived from the preceding one, $\mathbf{x}_i$, by means of a random move of one randomly chosen variable

$$x_{i+1}^m = x_i^m + \lambda_m \cdot \xi \qquad (1)$$

where ξ is a random quantity from the interval (−1, 1) and λ_m is a maximal allowed increment of the given variable. Whether the new state $\mathbf{x}_{i+1}$ is accepted or rejected is decided according to a transition probability function.[71,74]

This conventional Monte Carlo technique is not efficient because the conformational variables of peptide chains appear to be mutually dependent. When applied to proteins, eq 1, in which structural parameters are sequentially changed one after another, results in a very slow conformational motion.[41,43,49]

To model a ***simultaneous*** change of the structural parameters, we have used eq 2 (instead of the conventional (1))

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \lambda_m \cdot \xi \cdot \mathbf{g}^m \qquad (2)$$

where $\mathbf{g}^m$ is a randomly chosen ***basis vector*** from a basis $\mathbf{G} = \{\mathbf{g}^m\}$, ξ is a random quantity from an interval (−1, 1) and λ_m is a maximal acceptable move in direction $\mathbf{g}^m$. In this study λ_m was constant and equal to 10°. A basis **G** is formed from linear combinations of independent conformational parameters **x** for their change to be realized ***along the soft modes*** of the molecule. These modes are revealed by dispersion analysis of the distribution of the sequential configurational points of the Markov chain. Initially, basis **G** coincides with a unit matrix, and then the rotation of basis **G** is carried out after every 150 Monte Carlo steps such that basis vectors $\mathbf{g}^m$ are made equal to the eigenvectors of the covariant matrix formed from the previous 150 configurations **x**.[73]

2.2. Implementation of the VBMC Algorithm. VBMC was interfaced with the CHARMm force field.[75] Torsional distortion, van der Waals and electrostatic interactions, and hydrogen bonding were thus explicitly considered. The united atom approach was used for all carbons with nonpolar hydrogens. Point atomic charges of Weiner et al.[76] were used. A dielectric constant equal to the interatomic separation in angstroms was utilized to implicitly simulate solvation effects.[76] All calculations were performed on an 8-node IBM Scalable POWERparallel 2 [SP2] high performance computer at Queen's University. Each Monte Carlo simulated annealing computation required up to 120 h of CPU time.

Each run of the VBMC algorithm provides a unique final structure. This would represent the global energy minimum structure, regardless of the initial geometry, if the "cooling" was performed infinitely slowly. However, since this is impossible, several Monte Carlo simulated annealing computations must be performed from different starting points $\mathbf{x}_0$ using sufficiently high initial temperatures $\mathbf{T}_0$. An initial temperature of 1000 K is sufficient. The Markov chain is kept at a constant temperature $\mathbf{T}_0$ for 10^4 Monte Carlo steps (i.e., 0.7×10^6 Monte Carlo moves) before the initiation of "cooling". The temperature **T** is then decreased at every Monte Carlo move (2) by a small increment such that after 2×10^4 Monte Carlo steps (about 1.5×10^6 Monte Carlo moves) it reaches 120 K; at this time, a local molecular mechanical potential energy refinement is performed to identify the nearest stationary point on the potential energy surface.

2.3. Applying VBMC to NGF. The amino acid sequences of the termini of the neurotrophin molecules under study are presented in Figure 1. Residues are numbered from 1 to 118 for the first monomer and from 1′ to 118′ for the second monomer. The mutual orientation of the termini of the different monomers is determined by the disulfide bonds $\overline{\text{Cys}^{15}\ \text{Cys}^{80}}$ and $\overline{\text{Cys}^{68}\ \text{Cys}^{110'}}$ and hydrophobic interactions Val^{14}:$\text{Val}^{14'}$ and Val^{109}:$\text{Val}^{109'}$, which are absolutely conserved in all neurotrophins.[36] The conformational space of the peptide molecules consists of flexible torsional angles determining geometry of the amino terminus of one monomer and the carboxyl terminus of the other, while keeping the rest of the dimer fixed in the known geometry of mNGF.[36] Covalent bond lengths and bond angles of the termini are kept fixed. The geometry of each residue is determined only by flexible torsional angles of the peptide backbone (ϕ_i and ψ_i) and the side chain ($\chi_i1,\chi_i2,...$). All torsional angles of amino acid Pro except for ψ_i are fixed. Torsional angles describing structural motifs with the essentially planar geometry (such as peptide bonds or conjugated regions in side chains) are kept fixed. Likewise, the configurations of the sp^3-carbons are fixed. With these constraints, the total

(64) Lelj, F.; Grimaldi, P.; Cristianziano, P. L. *Biopolymers* **1991**, *31*, 663−670.
(65) Villani, V.; Tamburro, A. M. *J. Mol. Struct. (Theochem)* **1994**, *308*, 141−157.
(66) Kincaid, R. H.; Scheraga, H. A. *J. Comput. Chem.* **1982**, *3*, 525−547.
(67) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(68) Brooks III, C. L.; Case, D. A. *Chem. Rev.* **1993**, *93*, 2487−2502.
(69) Rao, M.; Pangali, C.; Berne, B. J. *Mol. Phys.* **1979**, *37*, 1773−1798.
(70) Rossky, P. J.; Doll, J. D.; Friedman, H. L. *J. Chem. Phys.* **1978**, *69*, 4628−4633.
(71) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(72) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Claredon Press: Oxford, 1987.
(73) Szentpály, L. v.; Shamovsky, I. L.; Nefedova, V. V.; Zubkus, V. E. *J. Mol. Struct. (Theochem)* **1994**, *308*, 125−140.
(74) Gunningham, G. W.; Meijer, P. H. E. *J. Comp. Phys.* **1976**, *20*, 50−63.
(75) (a) Brooks, B. R.; Bruccoleri, R. E.; Olason, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217. (b) The force field parameters are extracted from CHARMm release 23.1, Molecular Simulations, Inc.
(76) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

conformational space of the flexible termini of hNGF, mNGF, and BDNF consists of 74, 73, and 67 variables, respectively.

3. Results and Discussion

By diagonalizing the covariant matrix in the framework of the VBMC algorithm, rigid modes of a molecular system, which should slow down conformational changes, are supposed to be revealed simultaneously with soft ones. To test the efficiency of the VBMC algorithm, its performance was compared with the conventional Monte Carlo method at the equilibrium parts of the Markov chains generated at different constant temperatures. These comparisons showed that at high temperatures (700−1000 K) the basis variability results in decreasing acceptance rates. On the other hand, the temperature decrease reinforces the advantage of the VBMC approach, and at temperatures below 600 K acceptance rates of the VBMC Markov chains were significantly higher. Thus, at $T = 300$ K the acceptance rate of the conventional Markov chain was found to be 0.22 which is very close to that recommended by Kincaid and Scheraga as optimal (0.20).[66] The acceptance rate of the VBMC algorithm obtained at the same conditions was 0.26, i.e., 18% higher. These evaluations are in agreement with what one would expect. Indeed, at low temperatures only soft modes of molecules are involved in thermal motions, hence, they are especially populated in corresponding Monte Carlo generated conformational ensembles and can be revealed from them. Higher acceptance rates of the low temperature VBMC Markov chains over conventional ones mean that the average increase of the sizes of the Monte Carlo moves along soft modes is higher than their decrease along rigid modes. Thus, the VBMC algorithm not only changes directions of the Monte Carlo moves but also accelerates conformational changes by directing them along the soft modes of the molecule. Consequently, the VBMC algorithm is more efficient than conventional Monte Carlo method only at moderate and low temperatures, when thermal motions tend to take place along the valleys of the potential energy surface. This particular temperature region causes serious problems in simulated annealing algorithms, and an efficient algorithm for this region is especially needed.

Seven independent VBMC simulations for each of the six molecular systems under study were performed. Simultaneously, local energy refinements were periodically carried out from a number of points generated during the "slow cooling" stage. All local minimizations resulted in higher energy stationary points on the potential energy surface than those produced by VBMC computations. This means that the VBMC technique as an approach to global minimization cannot be replaced by relatively inexpensive periodic energy refinements from the high-temperature conformational space; at high temperatures only high-energy conformers are sufficiently populated. Although VBMC runs did not converge to the same conformation, the final structures differed insignificantly, i.e., the peptide backbone conformation remained almost unchanged, but there were limited differences in the side chain torsional angles. Multiple VBMC simulations identified rigid and flexible parts of the molecules. Minimum energy structures of all molecules are displayed in Figures 2−7.

3.1. Geometric Similarity of Bioactive NGF Congeners. Rigid regions of hNGF and mNGF which maintain the same conformation in all independent VBMC simulations are formed by residues 9-11 and 112′-118′, creating a richly hydrogen bonded (H-bonded) three-dimensional complex. Conformational variability within this rigid zone is primarily restricted to side chains and does not affect the geometry of the backbone. An interesting example of such side chain mobility occurs with

First monomer: amino terminus

	1				5					10					15
hNGF	S	S	S	H	P	I	F	H	R	G	E	F	S	V	C ...
mNGF	S	S	T	H	P	V	F	H	M	G	E	F	S	V	C ...
mNGFΔ3-9	-	-	-	-	-	-	-	S	S	G	E	F	S	V	C ...
hNGF-H4D	S	S	S	D	P	I	F	H	R	G	E	F	S	V	C ...
mNGFΔ1-8	-	-	-	-	-	-	-	-	M	G	E	F	S	V	C ...
BDNF	-	-	H	S	D	P	A	R	R	G	E	L	S	V	C ...

Second monomer: carboxyl terminus

					112'						118'
hNGF	... C	V	C	V	L	S	R	K	A	V	R
mNGF	... C	V	C	V	L	S	R	K	A	T	R
mNGFΔ3-9	... C	V	C	V	L	S	R	K	A	T	R
hNGF-H4D	... C	V	C	V	L	S	R	K	A	V	R
mNGFΔ1-8	... C	V	C	V	L	S	R	K	A	T	R
BDNF	... C	V	C	T	L	T	I	K	R	G	R

Figure 1. Amino acid sequences of the conformationally variable termini of the dimeric molecules under study. The residues are numbered from 1 to 118 for the first monomer and from 1′ to 118′ for the second monomer. The fixed parts of the proteins simulated by the X-ray geometry of mNGF are underlined.

$Arg^{114'}$, which is H-bonded to the carbonyl oxygen of the ninth amino acid of the amino terminus and which is therefore expected to have a fixed conformation. Nevertheless, $Arg^{114'}$ has two distinct conformers in which the H-bonding is realized by the hydrogen atoms attached to the different NH_2 groups, with the energy difference being about 1 kcal/mol. In addition to local conformational variability, hNGF and mNGF have a region of "cooperative" flexibility which is comprised of residues 1-8. Several distinct conformations exist within this region, with its conformational variability being caused by a number of possibilities for the H-bond network formed by residues 1-3.

Comparison of the most stable conformers of hNGF (Figure 2) and mNGF (Figure 3) reveals their striking conformational similarity in spite of several differences in the primary structure of the amino and carboxyl termini (Figure 1). There are only two noticeable differences in hNGF and mNGF, namely the side chains of $Arg^{114'}$ and $Arg^{118'}$ have slightly different conformations.

The essential difference between the amino acid sequences of the termini of hNGF and mNGF is the mM9hR substitution, which does not change the geometry of the molecule. Residues Met and Arg have different side chain types, i.e., the former is nonpolar while the latter is polar and positively charged under physiological conditions. It is surprising that this substitution does not affect geometry since it introduces the positive charge of Arg^9 which is located in an electrostatic field created by at least five neighboring charged functional groups. Conformational stability of hNGF is maintained by the exact fit of the Arg^9 side chain into a "hollow" in the rigid region, which is occupied by the Met^9 side chain in mNGF. Furthermore, the Arg^9 side chain, being in this "hollow", forms two additional interchain H-bonds with carbonyl oxygens of residues $Val^{117'}$ and $Arg^{118'}$. This further stabilizes the geometry of the hNGF termini.

The physical reason for the geometric similarity of the complexes of the termini of hNGF and mNGF is that the structural features of these two complexes are determined by similar H-bonding and electrostatic interactions. The major

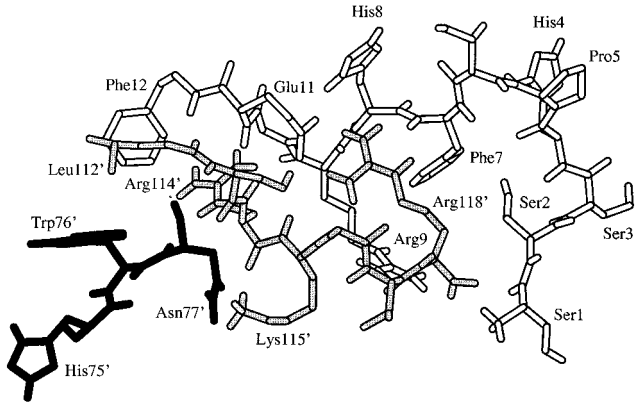

Figure 2. The structure of the complex of the amino and carboxyl termini of (1-118)hNGF homodimer. The carboxyl terminus is shown in grey, the amino terminus in white, and the conserved region $His^{75'}$–$Asn^{77'}$ of the "south" part of NGF in black. Only residues mentioned in the text are denoted.

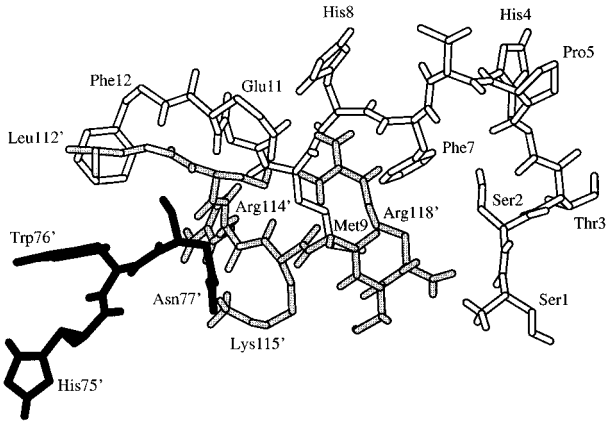

Figure 3. The structure of the complex of the amino and carboxyl termini of (1-118)mNGF homodimer.

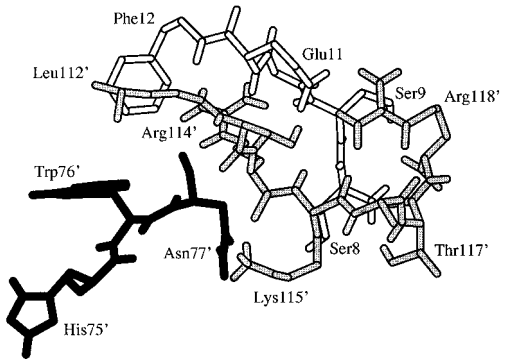

Figure 4. The structure of the complex of the amino and carboxyl termini of dimeric mNGFΔ3-9 molecule.

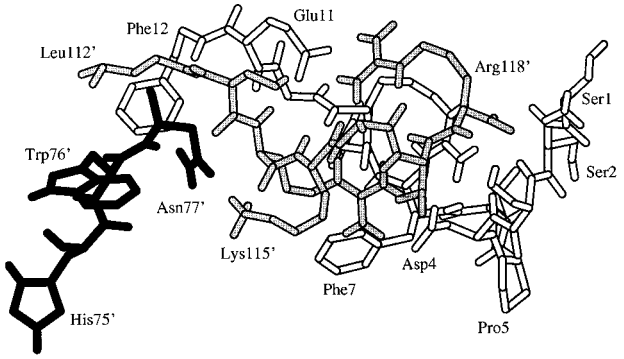

Figure 5. The structure of the complex of the amino and carboxyl termini of dimeric hNGF-H4D molecule.

stabilizing factor is the H-bonding and electrostatic interactions of charged side chains of residues Glu^{11} and $Arg^{118'}$. In addition, these side chains are involved in other H-bonding: the carboxyl oxygens of Glu^{11} are bonded to the δ-NH group of His^8, the hydroxyl group of $Ser^{113'}$ and the backbone NH group of $Lys^{115'}$; the ϵ-NH and η-NH_2 groups of $Arg^{118'}$ are bonded to the backbone Phe^7 and $Lys^{115'}$ carbonyl groups; side chains of residues 1−4 are involved in H-bonding with the backbone of the 1-8 loop. There are some other common interchain H-bonds in hNGF and mNGF. The ϵ-amino group of the $Lys^{115'}$ side chain forms a H-bond with the δ-carbonyl oxygen of the $Asn^{77'}$ residue, situated at the "south" part of NGF structure.[36]

In addition to H-bonding, the particular geometry of the termini in both molecules is also maintained by electrostatic interactions of positively charged residues, specifically His^4, His^8, $Arg^{114'}$, $Arg^{118'}$, and $His^{75'}$. An electrostatic repulsion between His^4 and boundary residues His^8 and $Arg^{118'}$ of the rigid region forces His^4 to be located as far as possible from the rigid region, separating the 1-8 loop from the rest of the structure. This tendency is further reinforced in hNGF by the presence of a third positively charged boundary residue Arg^9. The $His^{75'}$ amino acid located in the "south" part of NGF also participates in maintaining the structure of the complex. An electrostatic repulsion between $His^{75'}$ and $Arg^{114'}$ prevents H-bonding of the η-amino groups of $Arg^{114'}$ with residues $His^{75'}$, $Asn^{77'}$, and $Ser^{113'}$: indeed, removing the ϵ-proton from $His^{75'}$ results in these multiple H-bond contacts which change the global energy minimum structure of the complex.

The separation of flexible loop 1-8 from the rigid region is also determined by residue Pro^5. This amino acid imposes restrictions on conformational motion of the loop making it "bent" at this particular point and thereby defining the location of the adjacent His^4 residue at the most distant point from the rigid region. Thus, residues His^4 and Pro^5 are primarily responsible for the separation of the flexible loop from the rigid region.

Although mNGFΔ3-9 does not have the flexible loop, the energy optimized geometric features of its termini are very similar to those of the hNGF and mNGF rigid regions (Figures 2−4). There are only two structural elements which are specific for the rigid region of mNGFΔ3-9. First, although $Arg^{118'}$ has a different conformation, the result is functionally the same, namely both η-amino groups of the $Arg^{118'}$ side chain are H-bonded to the Glu^{11} side chain. In addition, the ϵ-NH group of $Arg^{118'}$, which is H-bonded to the Phe^7 carbonyl group of the flexible loop in both hNGF and mNGF, faces the carbonyl terminus in mNGFΔ3-9. Second, the "hollow", which is occupied by the ninth amino acid side chain in hNGF and mNGF, is now filled with the amino acid backbone of the remainder of the loop, i.e., residues 1−2. Functionally, the structure of the rigid region in mNGFΔ3-9 resembles that in hNGF because of the stabilizing interchain H-bond contacts of the moieties located in the "hollow". In addition, because of the positive charge of those moieties in hNGF and mNGFΔ3-9, the conformations of the charged $Arg^{114'}$ side chains in these molecules are very similar and slightly different from that of mNGF (Figures 2−4).

3.2. Geometric Distinction of Inactive Molecules. The energy minimum structures of the complexes formed by the amino and carboxyl termini of inactive molecules, hNGF-H4D, mNGFΔ1-8, and BDNF, do not possess the common structural features found in active molecules. Point mutation H4D in hNGF incorporates negative instead of positive charge into the 1-8 loop. As a result of this mutation, the loop is no longer

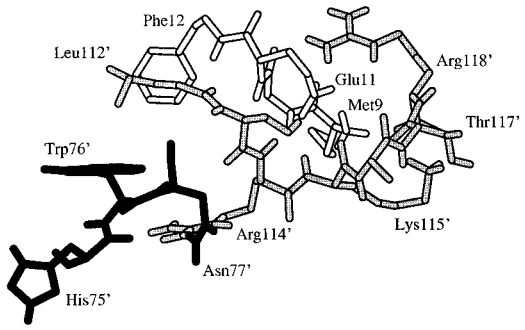

Figure 6. The structure of the complex of the amino and carboxyl termini of dimeric mNGFΔ1-8 molecule.

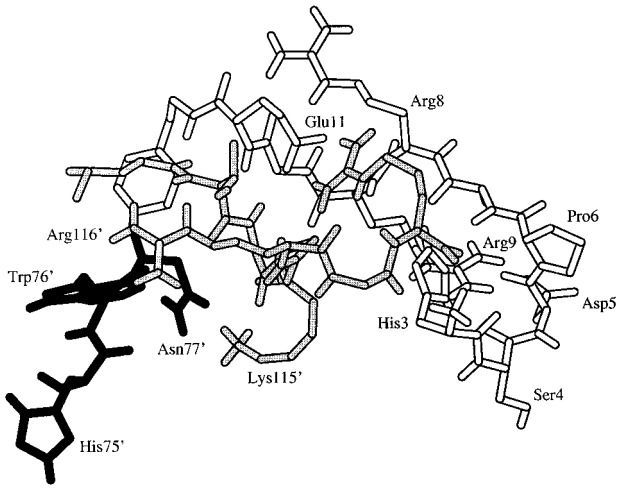

Figure 7. The structure of the complex of the amino and carboxyl termini of BDNF homodimer.

separated from the rigid region, forming a united complex instead (Figure 5). In this complex the negatively charged carboxyl group of Asp^4 is surrounded by four positively charged η-amino groups of Arg^9 and $Arg^{114'}$, and residue His^8 is not H-bonded to the Glu^{11} carboxyl group. Thus, because of the H4D point mutation, the common structural features inherent in hNGF and mNGF are lost.

Deletion 1-8 also changes dramatically the most stable structure of the rigid region of mNGF (Figure 6). The positively charged amino end is located between negatively charged carboxyl groups of Glu^{11} and $Arg^{118'}$, with charged side chains of $Arg^{114'}$, $Lys^{115'}$, and $Arg^{118'}$ being substantially shifted from their preferred positions in active molecules. Furthermore, the hydrophobic side chain of Met^9 is not packed in the "hollow" as in the active molecules, but rather is exposed to solvent. Thus, the electrostatic field caused by the positive charge of the amino end in mNGFΔ1-8, located in the vicinity of the carboxyl group of Glu^{11}, is responsible for the destruction of the geometry of the rigid region inherent in the active molecules.

The complex formed by the amino and carboxyl termini of BDNF is also significantly different from those formed by the active molecules (Figure 7). There are several reasons for the considerable conformational changes in BDNF. First, negatively charged residue Asp^5 eliminates the separation of the flexible loop from the rigid region and interacts with the positively charged His^3, Arg^9, and amino end. Second, because of its long tail, Arg^8 is unable to form the H-bond contact with the Glu^{11} side chain and forms bifurcated H-bonds with the carbonyl group of Gly^{10}. Third, $Arg^{116'}$ is H-bonded to $Glu^{66'}$ located in the "south" part of the molecule, which imposes a specific restriction on conformational possibilities of the carboxyl terminus of BDNF.

3.3. Explanation for the Observed Structure−Activity Relationships. The principal result of the molecular simulations presented is that the NGF molecules with full TrkA mediated activity considered in this study have regions with similar three-dimensional structure, whereas in all inactive analogues, the geometry of this region is altered mainly by novel electrostatic interactions. It is therefore reasonable to attribute the TrkA mediated activity to a specific conformation of this particular region. Experimental data consistent with this concept are presented below.

The VBMC computations show that the following amino acids are critical in determining the structure of the amino-carboxyl termini complex in wild type NGF dimers: His^4, Pro^5, His^8, Arg^9, Glu^{11}, $His^{75'}$, $Asn^{77'}$, $Lys^{115'}$, and $Arg^{118'}$. This is significant since residues His^4, Pro^5, Glu^{11}, $His^{75'}$, $Asn^{77'}$, and $Lys^{115'}$ are conserved in NGF molecules across all species.[36] The overlap between these two sets reflects a tendency to maintain the particular geometry of the complex in natural NGF molecules. In addition, as demonstrated, the substitution of the important residue Arg^9 by the essentially different Met^9 does not result in geometric changes of the complex, which is consistent with equivalent bioactivity of hNGF and mNGF.[29] Thus, because of multiple physical factors maintaining the geometry of the complex, some species variability observed in NGF termini[36] is acceptable and does not cause geometric changes in the rigid region.

It is shown that each NGF terminus has several important residues, which determine the particular structure of the complex. This is consistent with high sensitivity of biological activity of NGF to the primary structure of both the amino and carboxyl termini.[3,23,29,30,32−35] The lack of the TrkA mediated activity of the mNGFΔ3-11, mNGFΔ3-12, mNGFΔ3-13, mNGFΔ3-14, mNGFΔ9-14, and mNGFΔ112-118 deletion mutants[32] is thus explained by the absence of the key residues Glu^{11} and $Arg^{118'}$.

The separation of the flexible loop from the rest of the complex is crucial in permitting the rigid region to exist with definite geometric features. Residues His^4 and Pro^5 which cause this separation are absolutely conserved elements of the wild type NGF molecules derived from different species,[36] and their compulsory location in the middle of the flexible loop is well suited for facilitating the maximal separation. Point mutation P5A or H4D weakens or destroys the tendency of the loop to separate from the rigid region; accordingly, these mutations result in dramatic loss of TrkA related activity.[23]

The flexible loop is not an essential structural element for TrkA activation since the mNGFΔ3-9 deletion mutant, which does not contain the loop (Figure 4), is as active as wild type mNGF.[32] However, complete elimination of all loop residues results in almost inactive mutants mNGFΔ1-8[29,32,34] and hNGFΔ1-9.[23,29,30,31,35] Our molecular simulation studies enable the explanation for this apparent contradiction. It is demonstrated that the ***backbone*** of residues Ser^8-Ser^9 in the mNGFΔ3-9 mutant is naturally incorporated into the structure of the rigid region because it matches the geometry of the "hollow" occupied by either Met^9 or Arg^9 ***side chain*** in wild type NGF dimers. On the contrary, total truncation of the flexible loop in the mNGFΔ1-8 deletion mutant results in placing the positively

(77) (a) Kobe, B.; Deisenhofer, J. *Nature* **1993**, *366*, 751−756. (b) Kobe, B.; Deisenhofer, J. *Trends Biochem. Sci.* **1994**, *19*, 415−421. (c) Kobe, B.; Deisenhofer, J. *Nature* **1995**, *374*, 183−186.

(78) Yoder, M. D.; Keen, N. T.; Jurnak, F. *Science* **1993**, *260*, 1503−1507.

(79) Baumann, U.; Wu, S.; Flaherty, K. M.; McKay, D. B. *EMBO J.* **1993**, *12*, 3357−3364.

(80) Windisch, J. M.; Auer, B.; Marksteiner, R.; Lang, M. E.; Schneider, R. *FEBS Lett.* **1995**, *374*, 125−129.

charged amino end in the vicinity of the key H-bonded Glu^{11} and $Arg^{118'}$ charged side chains (Figure 3), which abolishes the desired geometry of the rigid region (Figure 6). The same applies to the hNGFΔ1-9 deletion mutant which has neither the 1-8 loop nor the stabilizing Arg^9 residue (Figure 2). This being the case, biological activity of amino terminus deletions of NGF should be determined by the length of the remainder of the terminus rather than its particular sequence. Indeed, available experimental data confirm this point. Deletion mutant mNGFΔ2-8 which has exactly the same number of residues in the amino terminus as the highly active mNGFΔ3-9 mutant is also very active.[32] Some decrease in biological activity of mNGFΔ2-8 is likely to be caused by unfavorable exposing of the Met^9 hydrophobic side chain to solvent, whereas in the mNGFΔ3-9 deletion mutant this side chain is replaced by the hydrophilic hydroxyl group (Figure 4). Deletion mutants mNGFΔ3-8[32] and hNGFΔ1-5,[23] which contain a greater number of amino terminus residues, are remarkably less active than wild type NGF because the remainder of the loop is too long to occupy the "hollow" but too short to separate from the rigid region. The present explanations for the observed structure−activity relationships enable the prediction that, for example, a S2G mutation in the mNGFΔ3-9 deletion mutant would retain high biological activity.

Although the bioactive conformation of the NGF binding determinant is predicted in this study, the question which NGF residues are involved in the NGF-TrkA recognition remains unclear. Most probably, those residues which are primarily responsible for maintaining the particular structure of the NGF termini do not directly interact with the receptor. Thus, side chains of His^4, Pro^5, His^8, Met^9, Arg^9, Glu^{11}, and $Arg^{118'}$ are unlikely to play a crucial role in molecular recognition. On the other hand, charged residues $Arg^{114'}$ and $Lys^{115'}$ which are conserved in NGF molecules across all species[36] but not involved in principal intramolecular electrostatic interactions are candidates for the centers of intermolecular electrostatic recognition. Although both $Arg^{114'}$ and $Lys^{115'}$ participate in stabilizing intramolecular H-bonding within the bioactive conformations of the NGF termini, their electrostatic properties as charged residues are not utilized.

The other structural element of the termini with potential for stereochemical fit to the TrkA receptor is a short exposed β-sheet motif located between α-carbons of residues 7 and 9 (Figures 2 and 3). On the one hand, H-bonding between β-pleated sheets is rather common for protein−protein interactions.[77−79] On the other hand, a short exposed β-sheet motif is generally inherent in the structure of a leucine-rich repeat[77] which has been recently identified as the TrkA recognition site for NGF.[24,80] Other regions of NGF might also be involved in the NGF-TrkA recognition.[23,37]

4. Conclusions

The Variable Basis Monte Carlo simulated annealing technique enables analysis of the low-energy conformational space of peptide molecules. Although multiple VBMC calculations do not consistently converge to the global energy minimum structure, they do permit the biologically active conformation to be evaluated. Definite physical factors maintain and distinguish the bioactive conformation of a molecule, and these factors can be understood by analyzing the low-energy conformational space. The precisely defined global energy minimum structure is unnecessary, especially considering that empirical force field functions are not ideal. In biologically relevant molecular modeling, exploring the bioactive conformation is more important than determining the global energy minimum. VBMC is a useful tool for facilitating this exploration.

The present VBMC simulations predict that the amino (1-11) and carboxyl (112′-118′) termini of dimeric (1-118)NGF form a complex comprised of a rigid region and a conformationally variable loop. The rigid region is formed by residues 9-11 and 112′-118′, while the variable loop consists of residues 1-8. The major stabilizing factor of the rigid region is the ionic contact of the Glu^{11} and $Arg^{118'}$ side chains. The separation of the loop from the rigid region arises from the electrostatic repulsion between His^4 and boundary residues of the rigid region His^8, Arg^9 (in hNGF), and $Arg^{118'}$. The geometry of the complex explains observed structure−activity relationships. It is the conformation of the rigid region which is seemingly recognized by the TrkA receptor, and therefore it represents the biologically active conformation of NGF for TrkA activation.

The split of the complex of the NGF termini into two distinct moieties could have important biological implications. Although only the structure of the rigid region is recognized by the TrkA receptor, the flexible loop is able to control conformational stability of the biologically active conformation of the rigid region and, consequently, the NGF-induced changes in TrkA function including receptor phosphorylation and subsequent biological events.

Acknowledgment. The support of Allelix Biopharmaceuticals Inc., Mississauga, Ontario is gratefully acknowledged. This work was supported in part by an operating grant from the National Sciences and Engineering Research Council of Canada. D.F.W. acknowledges an Ontario Ministry of Health Career Scientist Award. The authors thank the following: Dr. Tom Blundell for providing the NGF crystal solution coordinates; Dr. Magdalena Dory who identified a possible role for a specific interaction of the amino and carboxyl termini for biological activity of NGF; and Dr. Heather Gordon for helpful discussions.

Supporting Information Available: Table 1S presents the optimized conformational parameters of the amino and carboxyl termini of the considered NGF analogues (4 pages). See any current masthead page for ordering and Internet access instructions.

JA9611194