Research Article

Journal of PeptideScience

Received: 19 August 2011 Revised: 8 December 2011 Accepted: 12 December 2011 Published online in Wiley Online Library: 6 March 2012

(wileyonlinelibrary.com) DOI 10.1002/psc.2393

Hexapeptide fragment of carcinoembryonic antigen which acts as an agonist of heterogeneous ribonucleoprotein M

Nicholas Y. Palermo,[a] Peter Thomas,[b] Richard F. Murphy[a] and Sándor Lovas[a]*

Colorectal cancers with metastatic potential secrete the glycoprotein carcinoembryonic antigen (CEA). CEA has been implicated in colorectal cancer metastasis by inducing Kupffer cells to produce inflammatory cytokines which, in turn, make the hepatic micro-environment ideal for tumor cell implantation. CEA binds to the heterogeneous ribonucleoprotein M (hnRNP M) which acts as a cell surface receptor in Kupffer cells. The amino acid sequence in CEA, which binds the hnRNP M receptor, is Tyr-Pro-Glu-Leu-Pro-Lys. In this study, the structure of Ac-Tyr-Pro-Glu-Leu-Pro-Lys-NH_2 (YPELPK) was investigated using electronic circular dichroism, vibrational circular dichroism, and molecular dynamics simulations. The binding of the peptide to hnRNP M was also investigated using molecular docking calculations. The biological activity of YPELPK was studied using differentiated human THP-1 cells, which express hnRNP M on their surface and secrete IL-6 when stimulated by CEA. YPELPK forms a stable polyproline-II helix and stimulates IL-6 production of THP-1 cells at micromolar concentrations. Copyright © 2012 European Peptide Society and John Wiley & Sons, Ltd.

Supporting information may be found in the online version of this article.

Keywords: carcinoembryonic antigen; minimal sequence; molecular dynamics; replica-exchange; molecular docking; hnRNP M

Introduction

The glycoprotein carcinoembryonic antigen [1] (CEA) is one of a large family of 29 gene products. All these genes are members of the much larger immunoglobulin supergene family [2]. CEA has a molecular mass of 180–200 kD, depending on the extent of its glycosylation [2]. CEA is the prototypical tumor marker, and a large number of clinical studies have shown correlations between serum CEA levels and advanced colorectal cancer, in particular, in the presence of liver metastasis [3]. CEA has been implicated in the development of hepatic metastases from colorectal cancers on the basis of a direct relationship between CEA production and metastatic potential of human colorectal cancer cell lines [4,5]. Injection of CEA into mice prior to injection of weakly metastatic cancer cells increases liver metastasis [6]. Similarly, weakly metastatic colon cancer cell lines become highly metastatic when transfected with the cDNA coding for CEA [7,8]. CEA-producing colon cancers are also retained within the liver longer than non-CEA-producing cells, and this results in an increased metastatic potential for the liver [9]. This is consistent with the increase in adhesion to the endothelium being the key step in hepatic metastasis [10–13]. The cell involved in CEA clearance from the circulation is the liver-fixed macrophage or Kupffer cell [14]. CEA binding causes the activation of Kupffer cells, both *in vitro* [14] and *in vivo* [15]. This results in the secretion of inflammatory cytokines, which include IL-1-α, IL-1-β, IL-6, and IL-10, and tumor necrosis factor (TNF-α) [2,13]. This activation is the key to the role of CEA in liver metastasis.

CEA-induced signal transduction occurs through an 80-kD protein on the surface of Kupffer cells which is identical to the heterogeneous nuclear RNA-binding protein M (hnRNP M) [16]. hnRNP M is ubiquitous in mammalian tissues and is mostly localized in the nucleus, although it can also shuttle between the nucleus and cytoplasm carrying RNA [17,18]. hnRNP M is expressed as a surface protein in Kupffer cells but not in most other cell types [19]. The cellular distribution, level of expression, and relative amount of hnRNP M determine cellular specificity for CEA binding. hnRNP M acts as a cell surface receptor for proteins that contain the penta-peptide Pro-Glu-Leu-Pro-Lys [20]. The penta-peptide is found in CEA in the hinge region between the N-terminal immunoglobulin-like domain and the first immuno-

* *Correspondence to: Sándor Lovas, Department of Biomedical Sciences, Criss II, Room 313, Creighton University, 2500 California Plaza, Omaha, NE 68178, USA. E-mail: slovas@creighton.edu*

a *Departments of Biomedical Sciences, Creighton University, Omaha, NE, 68178, USA*

b *Department of Surgery, Creighton University, Omaha, NE, 68178, USA*

Abbreviations: *DIEA, N,N*-diisopropylethylamine; DMF, *N,N*-dimethylformamide; ELISA, Enzyme-linked immunosorbent assay; FBS, fetal bovine serum; HATU, 2-(1H-7-Azabenzotriazol-1-yl)–1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, *N*-hydroxybenzotriazole; NMP, *N*-methylpyrrolidinone; PMA, phorbol myristate acetate; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol; TIS, triisopropylsilane

Journal of PeptideScience

Scheme 1. Domain model of CEA. The hinge region in which the sequence YPELPK is found is marked by a box. The N-terminal domain is marked with N, the immunoglobulin-like domains are marked A1-B3, the C-terminal membrane anchor is marked with an arrow, glycosylation sites by black pins and disulfide bridges marked with double lines. Source: http://www.carcinoembryonic-antigen.de.

globulin loop domain (Scheme 1). The three-dimensional (3D) structure of the entire CEA has been determined by comparative modeling [21], and it further supports the notion that Pro-Glu-Leu-Pro-Lys is located in the above mentioned hinge region.

Patients with a mutation in the region of the CEA gene, which codes for this peptide, have extremely high circulating CEA levels, presumably because of the inability of Kupffer cells to clear the protein from the blood [22]. We, therefore, seek to determine the structure of the Pro-Glu-Leu-Pro-Lys–hnRNP M complex so as to design high-affinity inhibitors of this binding. Blocking the response of Kupffer cells to CEA may be a way to alter the hepatic micro-environment, making it less hospitable for tumor cell implantation and growth.

This investigation focuses on elucidating the structural features of Ac-Tyr-Pro-Glu-Leu-Pro-Lys-NH_2 (YPELPK) by using computational and circular dichroism methods. The 3D structure of the extracellular region of hnRNP M (residues 644–735) was determined by NMR spectroscopy (pdb id 2DGV) and was used as a model for investigating the interactions of YPELPK and hnRNP M *in silico*. ELISA was used to determine the biological activity of the sequence YPELPK by measuring the IL-6 production of human THP-1 cells exposed to YPELPK. The importance of each side chain of YPELPK was investigated using an Ala-scan, in which sequential residues of YPELPK were replaced by Ala (Table 1). The detailed analysis of the structure of YPELPK in free and bound states lays a foundation for the design of an antagonist of CEA binding and CEA-induced cancer metastasis.

Materials and Methods

Model Selection

The structure of Ac-Tyr-Pro-Glu-Leu-Pro-Lys-NH_2 (YPELPK) was selected as a model because in previous experiments [20], the N-terminal Tyr was included in the model peptide to facilitate radioiodonation for radioligand binding assays, but it is the residue that precedes the minimally required Pro-Glu-Leu-Pro-Lys in the sequence of CEA. Leaving Tyr as the N-terminal residue was also considered advantageous because aromatic residues are known to participate in a variety of weakly polar interactions, which could increase the binding of the peptide to hnRNP M [23,24]. The *N*-acetyl and amide-protecting groups were added to the N-terminus and C-terminus, respectively, in order to preserve the electronic structure of the backbone as in CEA.

Table 1. Ala-scan of YPELPK. Residues of YPELPK replaced with Ala are indicated in boldface fonts

Peptide	Sequence
YPELPK	Ac-Tyr-Pro-Glu-Leu-Pro-Lys-NH_2
A1	Ac-**Ala**-Pro-Glu-Leu-Pro-Lys-NH_2
A2	Ac-Tyr-**Ala**-Glu-Leu-Pro-Lys-NH_2
A3	Ac-Tyr-Pro-**Ala**-Leu-Pro-Lys-NH_2
A4	Ac-Tyr-Pro-Glu-**Ala**-Pro-Lys-NH_2
A5	Ac-Tyr-Pro-Glu-Leu-**Ala**-Lys-NH_2
A6	Ac-Tyr-Pro-Glu-Leu-Pro-**Ala**-NH_2

3D structures of both the N-terminal domain [25] and the positions of the C^{α} atoms of the entire CEA [26] have been published. The YPELPK sequence is omitted from the N-terminal domain structure and the C^{α}-trace of the whole CEA is of low resolution. Therefore, the structure of YPELPK must be determined *de novo*.

In preliminary docking experiments, Pro-Glu-Leu-Pro-Lys bound to residues 67–82 of hnRNP M. Therefore, a fragment of the protein, Ac-YMNGMKLSGREIDVRI-NH_2, which henceforth referred to as the CEA receptor fragment (CEARF), was synthesized to study its interaction with YPELPK with the use of electronic circular dichroism (ECD).

Binding Domain of hnRNP M

The structure of the extracellular domain of hnRNP M that binds CEA was obtained from the protein data bank [27] (pdb id 2dgv).

Peptide Synthesis

Ac-Tyr-Pro-Glu-Leu-Pro-Lys-NH_2 (YPELPK), its Ala-scan analogs, Ac-Pro-Glu-Leu-Pro-Lys-NH_2 (PELPK) and CEARF were synthesized at a 0.1 mmol scale using a CEM (CEM, Matthews, NC 28104 U.S.A.) Liberty microwave peptide synthesizer. A 5 molar excess of N^{α}-Fmoc-protected amino acids dissolved in NMP was mixed with a solution of 1.0 ml of 0.45 M HATU in DMF and 0.5 ml of 2 M DIEA in NMP at 75 °C and subjected to 25 W microwave irradiation for 10 min. We used 7.0 ml of 0.1 M HOBt dissolved in 20% (v/v) piperidine in DMF for the N^{α}-Fmoc removal at 75 °C and subjected it to 35 W microwave irradiation for 3 min.

Peptide-resins and a cleavage mixture containing 95% TFA, 2.5% H_2O, and 2.5% TIS (v/v/v) were stirred in a round bottom flask at 0 °C for 10 min and then at room temperature for 110 min. CEARF-resin was incubated in a mixture containing 86.5% TFA, 2.5% EDT 2.5% thioanisole, 5% phenol, 2.5% H_2O, and 1% TIS (v/v/v/v/v/v) with 3% NH_4I (w/w). Peptides were then precipitated with ice-cold ether, collected by filtration, dissolved in 10% AcOH in H_2O, and lyophilized.

Peptides were purified by RP-HPLC on a Phenomenex (Torrance, CA, USA) Luna C18 semi-preparative column (5 μm, 200 × 10.0 mm) using a gradient of 3%–60% organic phase over 60 min. The aqueous phase was the 0.1% aqueous TFA, and the organic phase was the 0.09% TFA in acetonitrile. All peptides were characterized using ESI-mass spectrometry.

Circular Dichroism Spectroscopy

The electronic circular dichroism (ECD) spectra were recorded using a Jasco (Jasco Inc., Easton, MD, USA) J-810 spectropolarimeter through performing 20 scans from 185 nm to 250 nm at a speed of 100 nm min^{-1} scan in a 0.05-cm path length quartz cell. The background spectra of the solvents were subtracted, and molar ellipticity was calculated using peptide concentrations determined by RP-HPLC [28] with a Jupiter C18 column (5 μm, 200 × 4.6 mm). The spectrum of 100 μM YPELPK was measured in a 15 mM NaH_2PO_4-Na_2HPO_4 (phosphate) buffer, pH 7.4, and in TFE at 4 °C. The spectra of 100 μM PELPK and CEARF were measured individually and in a 1 : 1 molar ratio mixture in phosphate

buffer. The addition spectra of PELPK and CEARF were compared with the spectrum of the mixture of the two peptides to determine if any binding event took place.

YPELPK was lyophilized three times in a solution of 0.1 M DCl in D_2O to remove TFA salts. The peptide was then dissolved at a concentration of 20 mg ml^{-1} in D_2O, and the vibrational circular dichroism (VCD) spectrum was recorded at 4 °C in a 75 μm path length CaF_2 cell using a BOMEM-Biotools (BOMEM-BioTools, Jupiter, FL, USA) Chiralir Fourier Transform VCD spectrometer. Data were collected for 4 h in one-hour blocks; each block consisted of 225 IR scans and 4500 VCD scans. The spectra were corrected by subtracting the spectrum of D_2O.

Molecular Dynamics Simulations

The GROMACS V3.3 [29] package was used for all molecular dynamics (MD) simulations. Replica-exchange molecular dynamics (REMD) [30] simulations were used to obtain the structure of YPELPK. Exchange temperatures for REMD simulations were obtained from a web-based (http://folding.bmc.uu.se/remd/) temperature predictor [31]. The exchange probability was 0.2, and the number of replicas was 24. The lower and upper temperature limits were 300 K and 450 K, respectively. Simulations were run for 100 ns in a $5 \times 5 \times 5$ nm rectangular box with periodic boundaries. The box containing the peptide was filled with TIP4P water [32]. All simulations used the OPLS-AA/L [33] force field. Long range electrostatic interactions and nonbonded interactions were treated by the twin-range method. The short range cutoff was 0.9 nm, and the long range cutoff was 1.4 nm. A reaction-field correction was used for long range electrostatic interactions, and a dispersion correction was used for energy.

The trajectory was demultiplexed using the TRJCAT program included in the GROMACS package. The demultiplexing matrix was generated using the demux_2007 perl script available from the GROMACS web site (http://www.gromacs.org). The structures of demultiplexed trajectory were assigned to clusters on the basis of the root-mean-square deviation (RMSD) of the backbone atoms with the GROMOS method [34]. The RMSD cutoff was 1.0 Å. Two sets of backbone atoms were used in the clustering analysis. The backbone atoms consist of all residues of YPELPK and the N-terminal and C-terminal protecting groups, and the residues PELPK-NH_2. These sets of backbone atoms were chosen in order to determine whether the inherent flexibility of the *N*-Ac-Tyr residue could mask the conformational stability of the peptide.

The MD simulation of the binding domain of hnRNP M with the best pose of YPELPK, as determined by the ligand-docking calculations, was run to determine the validity of the pose. The docked YPELPK–hnRNP M complex was obtained using the GLIDE [35] program (Glide, version 4.5, Schrödinger, LLC, New York, NY, 2007) as described in the ligand-docking section. The simulation was run for 100.1 ns at 300 K using the OPLS-AA/L force field and TIP4P water in an $8 \times 8 \times 8$ nm rectangular box. Long range electrostatic interactions and nonbonded interactions were treated with the twin-range method. The short range cutoff was 0.9 nm, and the long range cutoff was 1.4 nm. A reaction-field correction was used for long range electrostatic interactions, and a dispersion correction was used for energy. The first 0.1 ns of the trajectory was for system equilibration and it was excluded from the analysis.

Secondary structures were determined using the DSSP program [36]. Distances, dihedral angles, and hydrogen bonds were determined using the g_dist, g_chi, and g_hbond programs, respectively, which are included in the GROMACS package. Distances between residues were calculated using the center of the mass of the side chains. A hydrogen bond between YPELPK and hnRNP M was assigned when any donor group was within 3.5 Å of an acceptor group; the cutoff angle was 30°.

Essential dynamics analysis [37] was used to determine the effect of the presence of YPELPK on the internal motions of the hnRNP M. The g_covar program was used to generate mass-weighted covariance matrices of hnRNP M and of the YPELPK–hnRNP M complex on the basis of the positions of the C^{α} atoms. The set of atoms used to calculate the covariance matrix of the YPELPK–hnRNP M complex included only those atoms that corresponded to hnRNP M. The g_anaeig program was used to generate two-dimensional projections of the trajectories on the first and second eigenvectors.

The solvent accessible surface area (SASA) of the interface between hnRNP M and YPELPK was determined using the method described by Wassenaar and associates [38]. The g_sas program, which utilizes the double cubic lattice method [39], was used to calculate all SASA and interface surface area (ISA) values. The probe radius was 1.4 Å and any atom with an absolute charge less than −0.2 e was considered hydrophobic.

Ligand-Docking

The GLIDE program was used for all docking calculations in the high-precision XP mode. All ligands were regarded as flexible and were docked to a rigid receptor by using a proprietary version of the OPLS-AA force field. The initial structure of YPELPK was taken as the central structure of the largest cluster as determined by REMD. The positions of the conformers were then refined using the OPLS-2001 [32] force field. Ligand-docking scores were generated on the basis of these conformers by using a proprietary algorithm. The initial structures of the alanyl analogs of YPELPK were obtained by replacing the side chains of the residues of YPELPK using the YASARA package (http://www.yasara.org). All peptides were docked to the structure of hnRNP M (pbd id 2DGV). The bounding box for the calculations encompassed the entire structure of the receptor. Output was restricted to the best ten poses for each ligand. The affinities of the peptides were ranked in relation to that of the YPELPK, according to the GLIDE XP score which is equivalent to the ΔG of binding.

The interactions between the best ligand pose of YPELPK and hnRNP M were analyzed using YASARA. Distances of salt bridges and hydrogen bonds were measured using the positions of the H and O atoms. Distances involving aromatic rings were measured using a dummy atom in the geometric center of the aromatic ring as the reference point.

A weakly polar interaction was assigned when the distance between any amide bond, charged side chain, or aliphatic hydrogen was within 7.0 Å of the center of mass of an aromatic ring, and the angle formed between the vector of the interacting group and the plane of the aromatic ring was less than or equal to 45°.

ELISA

Human THP-1 cells from American Type Culture Collection (passage number 3; 80 000 cells per well) were placed into 24 well plates and incubated for 18 h at 37 °C in 0.5 ml RPMI-1640 medium with 10% FBS (v/v) and 200 ng ml^{-1} PMA. The medium was aspirated, and the cells were incubated for 4 h in RPMI-1640 medium containing 10^{-6}–10^{-9} M YPELPK. The Ala-scan analogs of YPELPK (Table 1) were also incubated at a concentration of

Journal of PeptideScience

10^{-6} M. The medium was removed and its IL-6 content was determined using an Invitrogen (Invitrogen, Carlsbad, CA, USA) Human IL-6 ELISA kit. Cell assays were in quadruplicate for each peptide, and the experiments were repeated six times. Absorbances at 450 and 550 nm were measured using a Tecan (Tecan Inc., Durham NC, USA) plate reader in multiple mode that used a 2 × 2 square pattern with three pulses per square. IL-6 levels were calculated as the difference between the indicator wavelength and the reference wavelength, 450 and 550 nm, respectively. The average absorbance of three empty wells were subtracted. Statistical significance of data was determined using ANOVA analysis and the Tukey Honestly Significant Difference (HSD) post-ANOVA method. The GNUMERIC (http://projects.gnome.org/gnumeric) program was used for statistical analysis.

Results

Circular Dichroism Spectroscopy

The ECD spectra of YPELPK in the phosphate buffer and TFE have single negative peaks at 201 and 198 nm, respectively (Figure 1). The ECD spectrum of YPELPK in the phosphate buffer has an additional positive peak at 215 nm, which is not present in the TFE. The VCD spectrum of YPELPK is a negative couplet with the negative band at 1618 cm^{-1} and a broad positive band at 1650 cm^{-1} (Figure 2). The IR spectrum of YPELPK contains peaks at 1515, 1620, 1640, and 1706 cm^{-1}.

The ECD spectra of PELPK and CEARF each have a single negative peak at 198 and 197 nm, respectively (Figure 3(A)). The spectrum of the mixture of PELPK and CEARF is more intense than the calculated addition spectrum of the two peptides; both spectra have a single negative peak at 197 nm (Figure 3(B)).

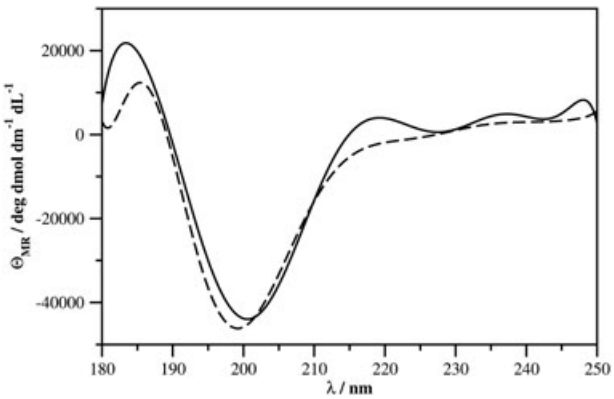

Figure 1. ECD spectra of YPELPK in phosphate buffer (black) and in TFE (dashed).

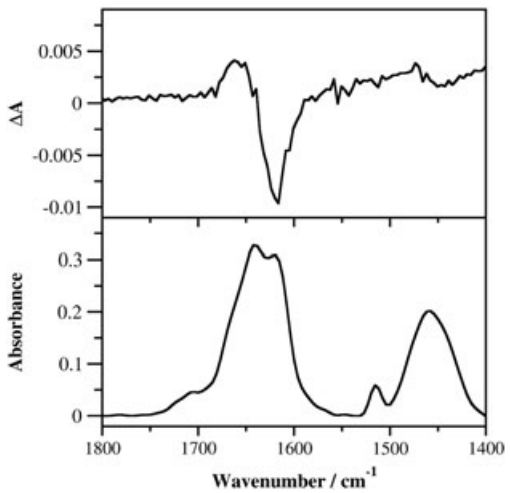

Figure 2. VCD (top) and IR (bottom) spectra of YPELPK in D_2O.

Molecular Dynamics Simulations

The demultiplexed replica-exchange MD simulation of YPELPK has a dominant backbone conformation consisting of 93.4% of the structures. This is not readily apparent until the flexible N-terminal Ac-Tyr residue is omitted from the clustering analysis (Table 2). The distribution of the backbone dihedral angles of YPELPK are shown in Figures 4 and S1 (see Supporting Information).

Most of the hydrogen bonds formed by YPELPK are with the last 26 residues of hnRNP M (L26) (Table 3, Figures 5 and S2). Cluster analysis of YPELPK reveals two major backbone conformations in the bound state, which consist of 53.9% and 29.4% of the structures (Table 4). This, again, was not apparent until the N-terminal Ac-Tyr residue was omitted from the clustering analysis. The first cluster is reflected to a change in the Pro2-Glu3 ψ dihedral angles, and the second cluster remains in the polyproline-II (PPII)-helical conformation (Figures 6 and S3).

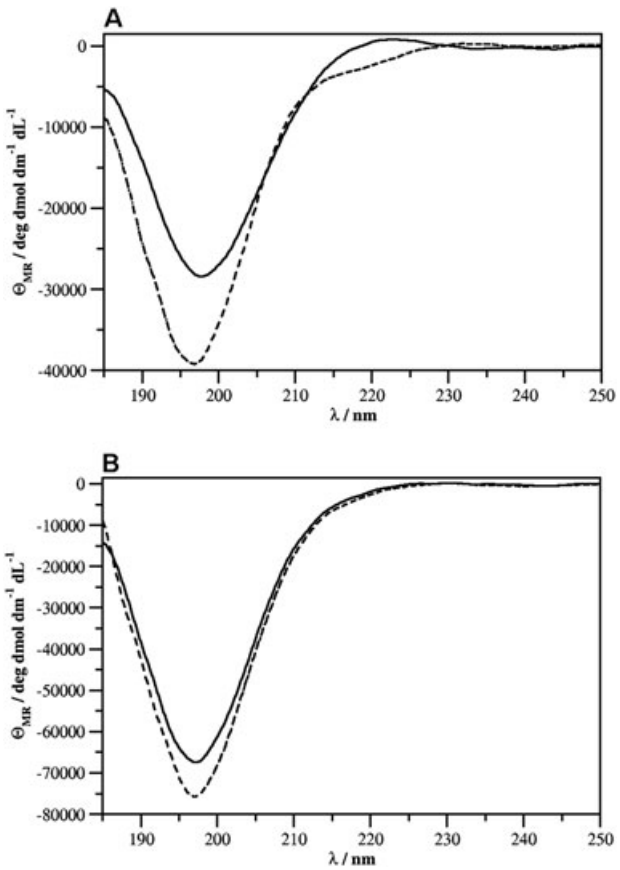

Figure 3. ECD spectra of PELPK and CEARF in phosphate buffer. (A) The spectrum of PELPK and CEARF are indicated in black and dashed lines, respectively. (B) The calculated addition spectrum and the measured spectrum of the mixture of the two peptides are indicated in black and dashed lines, respectively.

Table 2. Major clusters of the REMD simulation of the structure of YPELPK. The simulation was sampled every 20 ps for a total of 5000 structures

Sequence	Number of structures	% of structures
Ac-Tyr-Pro-Glu-Leu-Pro-Lys-NH_2	2531	50.6
Pro-Glu-Leu-Pro-Lys-NH_2	4670	93.4

Essential dynamics analysis shows that when the trajectory of the simulations of YPELPK–hnRNP M is projected on its first two eigenvectors, it samples less extreme regions than when sampling hnRNP M alone (Figure 7). The majority of the internal motions can be determined from the first two eigenvectors (Table 5).

Analysis of the SASA of the ligand-receptor complex reveals that L26 accounts for 77.6% of the total ISA (Table 6). The remainder of the interface is found in residues in the helical region immediately preceding the L26 and in the N-terminal coil.

Ligand-Docking

Analogs A1, A2, A3, and A6 have lower affinity for hnRNP M than has YPELPK; the difference in energy is less than 1.0 kcal mol^{-1} (Table 7). A4 and A5 have the least affinity for hnRNP M. The best ligand pose for YPELPK (Figure 8) is stabilized by a mix of two hydrogen bonds, two salt bridges, and four weakly polar interactions (Table 8).

ELISA

YPELPK at μM concentrations significantly increases IL-6 production by differentiated THP-1 cells (Figure 9(A)). The Ala-scan analogs of YPELPK activate THP-1 cells less than the parent peptide. A1 and A6, however, are still able to stimulate IL-6 production (Figure 9(B)).

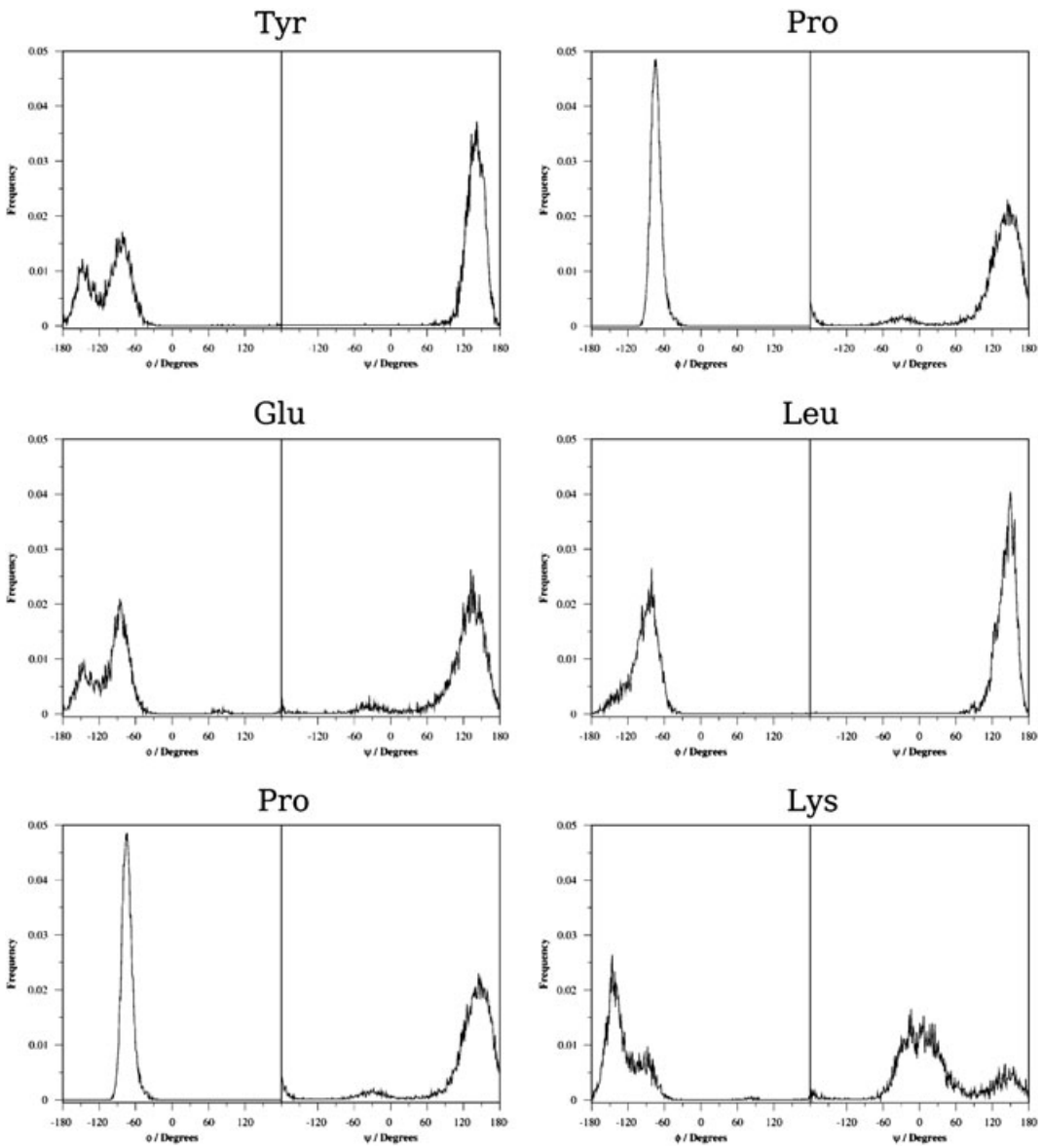

Figure 4. Distribution of ϕ and ψ torsional angles of YPELPK. Each profile depicts the relative frequency of each ϕ and ψ angle that occurred for each residue. The entire duration of the MD simulation was considered when calculating the frequencies.

Journal of PeptideScience

Table 3. Hydrogen bonds between YPELPK and hnRNP M. The simulation was sampled every 20 ps for a total of 5000 H-bond calculations

YPELPK–hnRNP M interaction site	Total number of H-bonds	% of H-bonds
Whole hnRNP M	13 319	100.00
L26	10 828	81.30

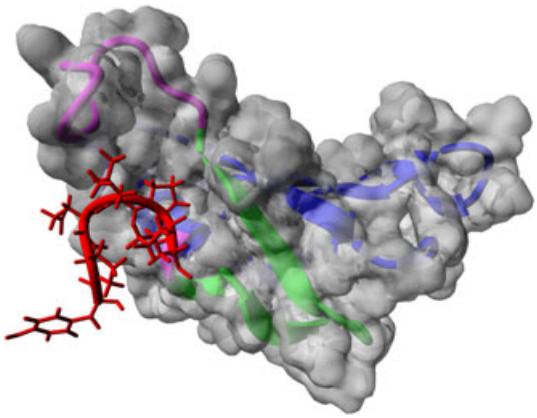

Figure 5. Snapshot of the MD trajectory of YPELPK (red) complexed with the C-terminal residues (magenta and green) of hnRNP M (van der Waals surface is in gray) at 80 ns. The backbone of the CEARF region of hnRNP M is in green; backbones of the non-interacting residues are blue.

Table 4. Major clusters of YPELPK from the simulation of the YPELPK–hnRNP M complex. The simulation was sampled every 20 ps for a total of 5000 structures

Sequence	Number of structures	% of structures
Ac-Tyr-Pro-Glu-Leu-Pro-Lys-NH_2	1104	22.1
Pro-Glu-Leu-Pro-Lys-NH_2	2694	53.9
Pro-Glu-Leu-Pro-Lys-NH_2	1469	29.4
Leu-Pro-Lys-NH_2	4910	98.2

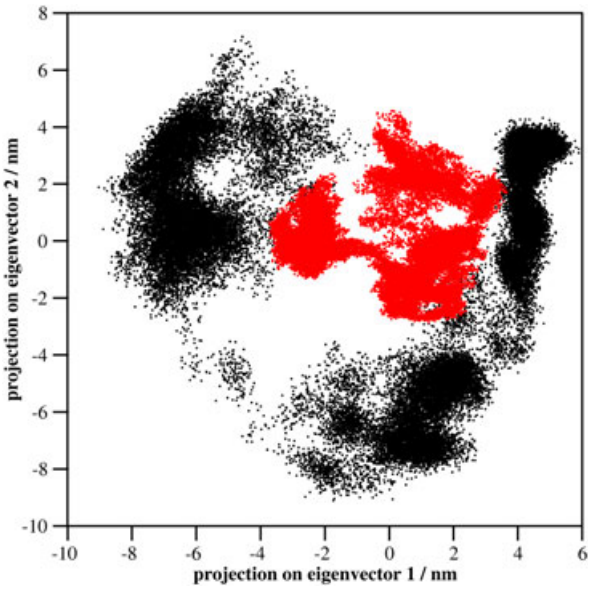

Figure 7. Projection of the trajectories of hnRNP M (black) and the YPELPK–hnRNP M complex (red) on their first two eigenvectors.

Discussion

The positive peak at 215 nm and negative peak at 201 nm in the ECD spectrum of YPELPK in phosphate buffer indicates either a random meander or a PPII-helical structure. The spectrum in TFE is blue shifted and its negative peak is moved to 198 nm and the positive peak at 215 nm disappears, indicating a less disordered structure. Short peptides such as YPELPK are typically flexible but YPELPK contains two prolyl residues which add rigidity to the backbone of the peptide, so the PPII helical structure is likely [40].

The negative peak at 198 nm in the ECD spectrum of PELPK in phosphate buffer indicates either a random meander or PPII helical structure, although the latter is most likely considering the proline content of the peptide. CEARF has a negative peak at 197 nm, which, considering the lack of Pro residues in CEARF, indicates a random meander or a β-hairpin structure. The

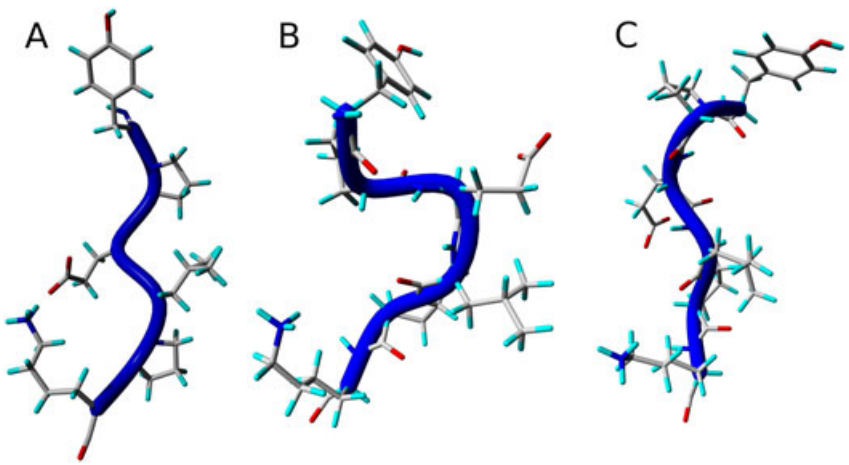

Figure 6. Conformations of YPELPK in free and bound states. (A) representative structure of the major cluster obtained from the analysis of the trajectory of the free YPELPK. (B) and (C) representative structures of the first and second major clusters, respectively, obtained from the analysis of the trajectory of the bound YPELPK.

Table 5. Eigenvalues and their sums of the first 10 eigenvectors of the essential movement of the YPELPK–hnRNP M simulations*

hnRNP M			YPELPK–hnRNP M complex		
Vector index	Eigenvalue / nm^2	Cumulative %	Vector index	Eigenvalue / nm^2	Cumulative %
1	21.56	36.28	1	3.61	42.04
2	13.00	59.83	2	1.73	62.20
3	6.75	71.18	3	0.81	71.66
4	3.46	77.00	4	0.64	79.11
5	2.33	80.92	5	0.36	83.28
6	1.75	83.86	6	0.21	85.75
7	1.20	85.88	7	0.18	87.81
8	1.04	87.62	8	0.17	89.80
9	0.84	89.03	9	0.09	90.89
10	0.68	90.17	10	0.09	91.88

* Eigenvalues with identical indexes do not represent the same motion because of the differences between the systems.

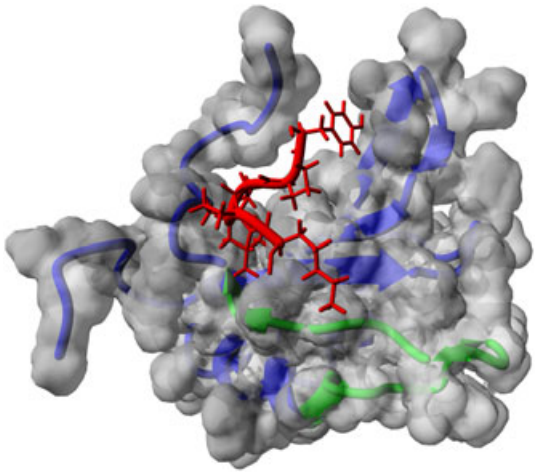

Figure 8. The best ligand pose of YPELPK (red) with hnRNP M (blue). The van der Waals surface of hnRNP M is in gray, and the backbone of the CEARF region of hnRNP M is green.

Table 6. Interface surface area components of hnRNP M. Only residues with ISA values greater than or equal to 0.01 nm^2 were considered in the analysis

Residue	ISA / nm^2	% ISA
Gly1	0.063	3.52
Ser2	0.088	4.97
Ser3	0.013	0.07
Ser6	0.014	0.08
Ala8	0.040	2.24
Glu62	0.061	3.45
Cys65	0.112	6.32
Arg66	0.211	11.87
Met67	0.010	0.06
Asn69	0.074	4.17
Asp 79	0.028	1.56
Val80	0.032	1.77
Arg81	0.150	8.42
Ile82	0.288	16.20
Arg84	0.302	16.98
Asn85	0.028	1.55
Pro89	0.050	2.81
Ser90	0.071	4.00
Ser91	0.101	5.70
Gly92	0.036	2.00
Total	1.773	99.62

Table 7. Relative free energy (ΔG_{rel}) of binding of YPELPK and its Ala-scan analogs. Energies of binding are relative to that of YPELPK. A positive value indicates less binding affinity

Peptide	ΔG_{rel} / kcal mol^{-1}
YPELPK	0.00
A1	0.80
A2	0.39
A3	0.75
A4	1.11
A5	3.33
A6	0.98

β-hairpin structure should be considered because the NMR structure of hnRNP M clearly shows that the residues corresponding to CEARF are in a β-hairpin conformation. The ECD spectrum of the 1 : 1 mixture of PELPK and CEARF in phosphate buffer has a negative peak at 197 nm and is more intense than the calculated addition spectrum of the two peptides. Therefore, the difference between the spectra indicates that PELPK and CEARF interact.

The VCD spectrum of YPELPK with a negative couplet centered at 1645 cm^{-1} indicates a PPII structure [41]. The peaks at 1515 and 1706 cm^{-1} in the IR spectrum are attributed to the motion of the tyrosyl ring and the CO stretch of the glutamic acid side chain, respectively [42]. The peak at 1620 cm^{-1} corresponds to a PPII structure [43], whereas the peak at 1640 cm^{-1} indicates an unordered structure. The center of the VCD couplet does not correspond to any one IR peak, probably, because the structure is a mix of PPII helix and random meander structure.

The molecular dynamics simulation indicates that YPELPK forms a stable PPII helix because the majority of the structures belong to a single cluster in which the peptide backbone is in the PPII helical conformation (Table 2). The *N*-acetyl group and tyrosyl residue of YPELPK have greater freedom of movement about the ϕ and ψ torsional angles than the rest of the residues of YPELPK (Figure S1). This torsional freedom introduced an artifact into the clustering analysis, which makes YPELPK appear to be a random coil structure. Although, when the backbone atoms of the *N*-acetyl group and the tyrosyl residue were omitted from the cluster analysis of the trajectory, YPELPK stayed in a single, stable conformation for 93.4% of the time.

The binding of YPELPK to the C-terminal region of hnRNP M, which does not associate with the cell membrane, indicates that no steric constraints would be imposed on binding. The weakly polar interactions between YPELPK and hnRNP M consist of two cation-π interactions, a CH-π interaction, and an aromatic-amide interaction (Table 7). The smaller loss of binding energy, 0.80 kcal mol^{-1}, when the Tyr side chain in A1 was replaced than when the Leu and Pro side chains were replaced in A4 and A5, respectively, was unexpected. The Tyr residue participates in a hydrogen bond and two weakly polar interactions, whereas the Leu and Pro residues in positions 4 and 5 do not have any direct interactions with the receptor. Even replacing the charged side chains, as in

Journal of PeptideScience

Table 8. YPELPK–hnRNP M interactions. Distances involving aromatic rings were measured using a dummy atom in the geometric center of the aromatic ring for reference

Functional group of YPELPK	Functional group of hnRNP M	Interaction type	Distance / Å
Tyr1 HO	Glu43 O^{ε}	H-bond	2.10
Tyr1 aromatic ring	Lys48 $^{\varepsilon}$HN	Weakly polar	3.44
Tyr1 aromatic ring	Lys41 $^{\varepsilon}$HN	Weakly polar	5.95
Pro2 C^{β}	Phe12 aromatic ring	Weakly polar	3.16
Glu3 O^{ε}	Lys41 $^{\varepsilon}$HN	Salt bridge	1.71
Glu3-Leu4 peptide bond	Phe12 aromatic ring	Weakly polar	6.15
Lys6 $^{\varepsilon}$HN	Asp79 O^{δ}	Salt bridge	1.81
C-terminal NH_2	Arg81 $^{\eta}$HN	H-bond	2.28

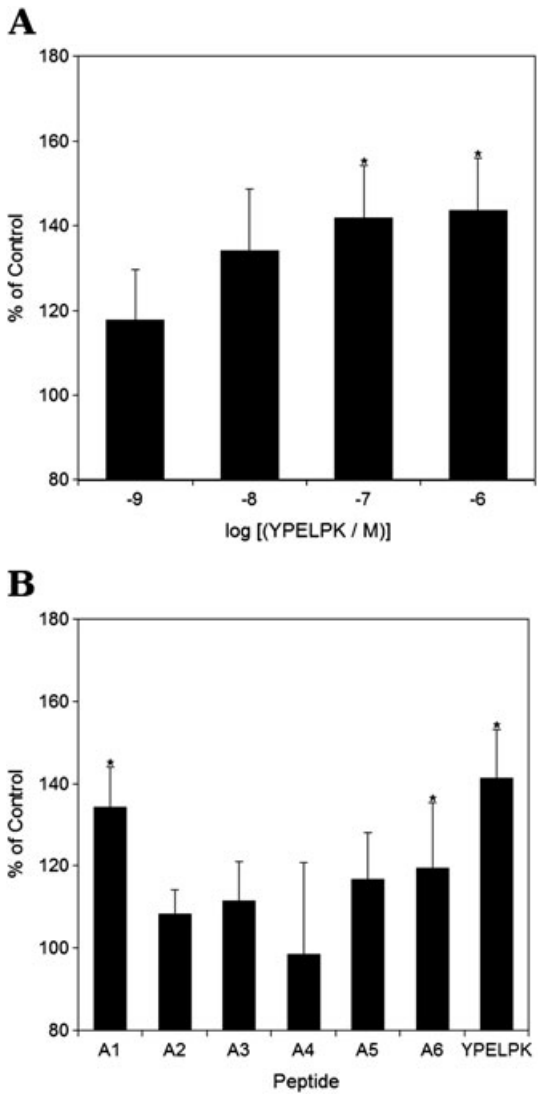

Figure 9. IL-6 production of THP-1 cells treated with YPELPK and analogs. (A) YPELPK dose response; (B) YPELPK and Ala-scan analogs. *, $p < 0.05$.

peptides A2 and A6, causes less loss of binding energy, 0.39 and 0.98 kcal mol^{-1}, respectively, than replacing the Leu and Pro residues. Thus, the conformational rigidity given by the Leu and Pro residues seems to be more important in binding than that given by other side chains, although these may also participate in electrostatic or weakly polar interactions.

MD simulations based on the ligand pose generated by GLIDE suggest, that after 10 ns, YPELPK shifts from the GLIDE-predicted binding site to a new site while in the complex with hnRNP M on the L26, and it remains there through the remainder of the simulation (Figure 5). This is supported by the SASA analysis. This indicates that the L26 is the major component of the interface of hnRNP M which is in contact with YPELPK (Table 5). The L26 as the binding site of YPELPK is further validated by the ECD results for the PELPK-CEARF mixture, as the two peptides clearly interact, and the sequence of CEARF is in a subset of L26.

The torsional freedoms of the two residues, Pro2 and Glu3, in bound YPELPK are higher than in the free peptide (Figure S3), suggesting induced fit binding. The increased torsional freedom of bound YPELPK is also due in part to the interaction of Glu and Lys residues with the receptor rather than to the formation of a salt bridge between them (Figure 5). Although the freedom of movement of bound YPELPK is increased, the internal movement of hnRNP M in the complex with YPELPK is diminished, as shown by the essential dynamics analysis. This could indicate the formation of a stable ligand-receptor complex. These data agree with a database analysis conducted by London and associates [44] which shows that receptors do not undergo significant conformational change, but the peptide ligand, which requires flexibility to fit into the binding site, does. The maintenance of the same conformation by the last three residues in 98.2% of the structure of the trajectory of bound YPELPK, even though the peptide is more flexible in the bound state, confirms the importance of these residues in binding, as predicted by GLIDE.

The ELISA shows that although CEA stimulates IL-6 production by THP-1 cells at a concentration as low as 30 pM [16], YPELPK still does so at μM concentrations. The high-resolution 3D structure of CEA is not available yet, so it could not be docked to the hnRNP M. CEA may have higher affinity for hnRNP M than YPELPK because it has more residues available for interaction. The slight production of IL-6 by A1 and A6 may reflect stronger binding than by A4 and A5.

In summary, YPELPK has a well-defined PPII structure in its unbound state and is biologically active at μM concentrations. It is remarkable that such a short sequence YPELPK can mimic the biological effect of the large CEA protein. In previous studies of the biological activity of the YPELPK sequence [20], the peptide was conjugated to albumin. It has now been shown that YPELPK alone is biologically active at concentrations low enough to be used as a model to develop an antagonist of CEA at the hnRNP M receptor.

Acknowledgement

This work was supported by the NIH-INBRE grant P20 RR016469.

References

1 Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. *J. Exp. Med.* 1965; **122**: 467–481.
2 Hammerstrom S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin. Cancer Biol.* 1999; **9**: 67–81.
3 Thomas P, Toth CA, Saini KS, Jessup JM, Steele GD. The structure, metabolism and function of the carcinoembryonic antigen gene family. *Biochim. Biophys. Acta* 1990; **1032**: 177–189.
4 Thomas P, Zamcheck N. Role of the liver in clearance and excretion of circulating carcinoembryonic antigen (CEA). *Dig. Dis. Sci.* 1983; **28**: 216–224.
5 Wagner HE, Toth CA, Steele GD, Thomas P. Metastatic potential of human colon cancer cell lines: relationship to cellular differentiation and carcinoembryonic antigen production. *Clin. Exp. Metastasis* 1992; **10**: 25–31.
6 Tibbetts LM, Doremus CM, Tzanakakis GN, Vezeridis MP. Liver metastases with 10 human colon carcinoma cell lines in nude mice and association with carcinoembryonic antigen production. *Cancer* 1993; **71**: 315–321.
7 Hostetter RB, Augustus LB, Mankarious R, Chi K, Fan D, Toth CA, Thomas P, Jessup JM. Carcinoembryonic antigen as a selective enhancer of colorectal cancer metastasis. *J. Natl. Cancer Inst.* 1990; **82**: 380–385.
8 Thomas P, Gangopadhyay A, Steele G, Andrews C, Nakazato H, Oikawa S, Jessup JM. The effect of transfection of the CEA gene on the metastatic behavior of the human colorectal cancer cell line MIP-101. *Cancer Lett.* 1995; **92**: 59–66.
9 Hashino J, Fukada Y, Oikawa S, Nakazato H, Nakanishi T. Metastatic potential of human colorectal carcinoma SW1222 cells transfected with cDNA encoding carcinoembryonic antigen. *Clin. Exp. Metastasis* 1994; **12**: 324–328.
10 Jessup JM, Petrick AT, Toth CA, Ford R, Meterissian S, O'Hara CJ, Steele G, Thomas P Carcinoembryonic antigen: enhancement of liver colonisation through retention of human colorectal carcinoma cells. *Br. J. Cancer* 1993; **67**: 464–470.
11 Gangopadhyay A, Lazure DA, Thomas P. Adhesion of colorectal carcinoma cells to the endothelium is mediated by cytokines from CEA stimulated Kupffer cells. *Clin. Exp. Metastasis* 1998; **16**: 703–712.
12 Langley RR, Fidler IJT. Tumor cell-organ microenvironment interactions in the pathogenesis of cancer metastasis. *Endocrine Revs.* 2007; **28**: 297–321.
13 Gangopadhyay A, Bajenova O, Kelly TM, Thomas P. Carcinoembryonic antigen induces cytokine expression in Kupffer cells: implications for hepatic metastasis from colorectal cancer. *Cancer Res.* 1996; **56**: 4805–4810.
14 Toth CA, Thomas P, Broitman SA, Zamcheck N. Receptor-mediated endocytosis of carcinoembryonic antigen by rat liver Kupffer cells. *Cancer Res.* 1985; **45**: 392–397.
15 Edmiston K, Gangopadhyay A, Shoji Y, Nachman A, Thomas P, Jessup JM. *In vivo* induction of murine cytokine production by carcinoembryonic antigen. *Cancer Res.* 1997; **57**: 4432–4436.
16 Bajenova OV, Zimmer R, Stolper E, Salisbury-Rowswell J, Nanji A, Thomas P. Heterogeneous RNA-binding protein M4 is a receptor for carcinoembryonic antigen in Kupffer cells. *J. Biol. Chem.* 2001; **276**: 31067–31073.
17 Kim JH, Hahm B, Kim YK, Choi M, Jang SK. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* 2000; **298**: 395–405.
18 Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* 2002; **3**: 195–205.
19 Bajenova O, Stolper E, Gapon S, Sundina N, Zimmer R, Thomas P. Surface expression of heterogeneous nuclear RNA binding protein M4 on Kupffer cell relates to its function as a carcinoembryonic antigen receptor. *Exp. Cell Res.* 2003; **291**: 228–241.
20 Gangopadhyay A, Thomas P. Processing of carcinoembryonic antigen by Kupffer cells: recognition of a penta-peptide sequence. *Arch. Biochem. Biophys.* 1996; **334**: 151–157.
21 Bates PA, Luo J, Sternberg JE. A predicted three-dimensional structure for the carcinoembryonic antigen (CEA). *FEBS Lett.* 1992; **301**: 207–214.
22 Zimmer R, Thomas P. Mutations in the carcinoembryonic antigen gene in colorectal cancer patients: implications on liver metastasis. *Cancer Res.* 2001; **61**: 2822–2826.
23 Palermo NY, Csontos J, Murphy RF, Lovas S. Role of aromatic residues in stabilizing the secondary and tertiary structure of avian pancreatic polypeptide. *Int. J. Quant. Chem.* 2008; **108**: 814–819.
24 Hatfield MPD, Palermo NY, Csontos J, Murphy RF, Lovas S. Quantum chemical quantification of weakly polar interaction energies in the TC5b miniprotein. *J. Phys. Chem. B* 2008; **112**: 3503–3508.
25 Korotkova N, Yang Y, Le Trong I, Cota E, Demeler B, Marchant J, Thomas WE, Stenkamp RE, Moseley SL, Matthews S. Binding of Dr adhesins of *Escherichia coli* to carcinoembryonic antigen triggers receptor dissociation. *Mol. Microbiol.* 2008; **67**: 420–434.
26 Boehm MK, Perkins SJ. Structural models for carcinoembryonic antigen and its complex with the single-chain Fv antibody molecule MFE23. *FEBS Lett.* 2000; **475**: 11–16
27 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res.* 2000; **28**: 235–242.
28 Szendrei GI, Fabian H, Mantsch HH, Lovas S, Nyeki O, Schon I, Otvos L, Jr. Aspartate-bond isomerization affects the major conformations of synthetic peptides. *Eur. J. Biochem.* 1994; **226**: 917–924.
29 van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 2005; **26**: 1701–1718.
30 Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 1999; **314**: 141–151.
31 Patriksson A, van der Spoel D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* 2008; **10**: 2073–2077.
32 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983; **79**: 926–935.
33 Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 2001; **105**: 6474–6487.
34 Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed Engl.* 1999; **38**: 236–240.
35 *Glide, Version 4.5*. Schrödinger, LLC: New York, NY, 2007.
36 Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983; **22**: 2577–2637.
37 de Groot BL, Amadei A, van Aalten DMF, Berendsen HJC. Towards an exhaustive sampling of the configurational spaces of the two forms of the peptide hormone guanylin. *J. Biomol. Struct. Dynam.* 1996; **13**: 741–751.
38 Wassenaar TA, Quax WJ, Mark AE. The conformation of the extracellular binding domain of Death Receptor 5 in the presence and absence of the activating ligand TRAIL: a molecular dynamics study. *Proteins* 2008; **70**: 333–343.
39 Eisenhaber F, Lijnzaad P, Argos, P, Sander C, Scharf M. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* 1995; **16**: 273–284.
40 Rucker AL, Creamer TP. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci.* 2002; **11**: 980–985.
41 Dukor RK, Keiderling TA. Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides. *Biopolymers* 1991; **31**: 1747–1761.
42 Jung C. Insight into protein structure and protein-ligand recognition by Fourier transform infrared spectroscopy. *J. Mol. Recognit.* 2000; **13**: 325–351.
43 Dukor RK, Keiderling TA. Mutarotation studies of poly-L-proline using FTIR, electronic and vibrational circular dichroism. *Biospectroscopy* 1996; **2**: 83–100.
44 London N, Movshovitz-Attias D, Schueler-Furman O. The structural basis of peptide-protein binding strategies. *Structure* 2010; **18**: 188–199.

