Molecular recognition in gas phase: theoretical andexperimental study of non-covalent protein-ligand complexes by mass-spectrometry

Andrey Dyachenko

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Programa de Quìmica Orgànica

Tesi Doctoral

Molecular recognition in gas phase: theoretical and experimental study of non-covalent protein-ligand complexes by mass-spectrometry

Andrey Dyachenko

Dirigida i revisada per:

Dr. Ernest Giralt
(Institut de Recerca Biomèdica Barcelona
Universitat de Barcelona)

Barcelona, 2013

Tesi Doctoral

Molecular recognition in gas phase: theoretical and experimental study of non-covalent protein-ligand complexes by mass-spectrometry

Andrey Dyachenko

Departament de Quìmica Orgánica

Facultat de Quìmica

Univesitat de Barcelona

2013

Index

Symbols and abbreviations

Symbols:

$[L]$ ($[S]$)	concentration of ligand (substrate)
$[M]_P$	parent ion peak intensity
$[P]$ ($[E]$)	concentration of protein (enzyme)
B_0	static magnetic field
C_N	total population of protein species with N ligand molecules bound specifically
$D, D_x, D_y, D_z, D_\perp, D_\parallel$	rotational diffusion tensor and its components
E	internal energy
E_A	activation energy of unimolecular dissociation
h	Planck constant
H	enthalpy
$J(\omega)$	spectral density
K_{COM}	kinetic energy in center-of-mass frame
K_D	dissociation constant
K_i	binding constant of i-th lignad
Kn	non-specific binding constant,
m_I, m_G	mass of ion and gas molecule, respectively
$P(E,K_{COM},\Delta E)$	probability density function
q	elementary charge,
R_1	longitudinal relaxation rate
R_2	transverse relaxation rate
R_{ex}	chemical exchange rate constant
r_{NH}	NH bon
S	model-free order parameter
S	entropy
S_s, S_f	“slow” and ”fast” model-free order parameters
T	temperature
T_1	longitudinal relaxation time
T_2	transverse relaxation time
V, V_C, V_{COL}:	collisional voltage
v_I, v_G	velocity of ion and gas molecule, respectively
Y	fractional binding site saturation
z	charge state
α	maximum number of different bound states visible in the spectrum

γ_H, γ_N	gyromagnetic ratios of ^1H and ^{15}N
ΔE	internal energy step
ΔG	Gibbs free energy
$\Delta\delta$	chemical shift displacement
$\Delta\sigma$	chemical shift anisotropy
ε	dielectric permittivity
ε_0	dielectric permittivity of vacuum
η	energy transfer efficiency
μ_0	magnetic constant,
τ_e	model-free correlation time
τ_s, τ_f	"slow" and "fast" model-free correlation times
ω_H, ω_N	Larnour frequencies of ^1H and ^{15}N

Abbreviations:

Amino acids:

Ser	S	serine
Asn	N	asparagine
Thr	T	threonine
Tyr	Y	tyrosine
Gln	Q	glutamine
Cys	C	cysteine
Trp	W	triptophane
Glu	E	glutamic acid
Met	M	metionine
Asp	D	aspartic acid
Gly	G	glycine
His	H	histidine
Ile	I	isoleucine
Phe	F	phenilalanine
Arg	R	arginine

AmAc	ammonium acetate
APCI	atmospheric pressure chemical ionisation
ATP	adenosine triphosphate
BIRD	blackbody infrared dissociation
CID	collision induced dissociation

COM	center of mass
CSA	chemical shift perturbation
ECD	electron capture dissociation
EDAA	ethylenediammonium acetate
EMF	extended model-free
ESI	electrospray ionization
ESSI	electro-sonic spray ionization
ETD	electron transfer ionization
FT-ICR	Fourier-transform ion cyclotron resonance
HCD	high-energy collision dissociation
HSQC	heteronuclear single quantum coherence
IDP	Intrinsically disordered proteins
IMS	ion mobility spectrometry
IRMPD	infrared multiphoton dissociation
ITC	isothermal titration calorimetry
KNF	Koshland-Nemethy-Filmer cooperativity model
MALDI	matrix-assisted laser desorption-ionization
MF	model-free
MS	mass-spectrometry
MWC	Monod-Wyman-Changeux cooperativity model
NMA	normal mode analysis
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PGF	placental growth factor
PLI	protein-ligand interaction
PPI	protein-protein interactions
RMSD	root-mean-square deviation
RSDM	reduced spectral density mapping
SID	surface induced dissociation
SORI	sustained off-resonance irradiation
TEAA	triethylammonium acetate
ToF	time-of-flight
TWIG	traveling wave ion guide
VEGF	vascular endothelial growth factor

1.

Introduction

1.1 Theory of non-covalent interactions

1.1.1 Non-covalent interactions in living organisms.

Biological molecules in living organisms fulfill their functions mainly through weak reversible interactions between each other. Signal transduction, DNA/RNA replication, other enzymatic processes and virtually any other process that underlies vital functions of amoeba and humans alike requires that molecules “talk” to each other [1], [2]. The “language” they speak is a language of non-covalent interactions.

Building of complete protein interactions map, or “interactome” of the entire cell is difficult task that has not been completed yet even for the simplest organisms. The building of such a map for the entire human body is sometimes referred to as “Human Interactome Project” [3][4], in analogy with Human Genome Project, completed in 2003. This analogy is not accidental but rather reflects the fact that the task of building human protein interaction map is at least as difficult and demanding as human genome project was back in 1990 [5]. To put this statement in perspective, the size of human proteome is estimated to be around 650000 binary interactions [6], of which less than 0.3% are identified today [7].

On the other hand, deciphering the mechanism of every single protein-protein or protein-ligand interaction is not a trivial task. Biological protein-protein and protein-ligand interactions (PPI and PLI) are generally characterized by two parameters: affinity, or the energy of binding between interacting partners; and specificity, or the ability of the protein to recognize its partner (or partners) and interact selectively with it (or them). With more than 30000 protein-coding genes identified in humans, proteins are operating in an extremely crowded environment [8]. The ability to distinguish correct interaction partner from thousands of other molecules requires extreme levels of specificity, that is achieved by a fine tuning of interaction interfaces in terms of placing right chemical groups at correct spatial points and allowing them to move in a correct fashion.

In addition to the specificity, each binary interaction is characterized by its affinity. The energy released upon binding of two molecules together determines the strength of this interaction. This energy can range between 10 and 80 kJ/mol, that corresponds to binding constants of 10^2 to 10^{14} M^{-1} [9], [10]. This energy is determined by the sum of entropic and enthalpic component (eq. 1.1) [11].

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

where ΔG in this case is the Gibbs free energy of binding, ΔH and ΔS *are* the enthalpy and the entropy of binding respectively and T is temperature. The processes that occur spontaneously are always associated with the decrease of the Gibbs free energy ($\Delta\Delta G < 0$). In case of non-covalent interactions that means that the complex formation has to lead to either release of heat ($\Delta\Delta H < 0$) or increase of entropy ($\Delta\Delta S > 0$) or both. Moreover, the balance between two terms has to be negative. For example, entropy penalty for putting two molecules together should be less than enthalpic gain for bond formation, otherwise binding will not occur [12]. The process of finding an interaction partner and binding to it with determined affinity is referred to as molecular recognition [13].

1.1.2 Thermodynamics of non-covalent interactions

When the complex between two proteins or between a protein and a ligand is formed, the system experiences a decrease in rotational and translational entropy $\Delta G_{T,R} = T\Delta S_{T,R}$. Although binding severely restricts the relative motions of the interaction partners, they most probably do not behave completely like a rigid body. In case of full restriction of relative motion the entropic penalty would be so large it would not allow binding to take place [14]. This suggests that binding partners retain some degree of flexibility relative to each other. In addition to the overall translational and rotational entropy loss, molecules often lose internal degrees of freedom due to the internal rotors that become restricted upon binding. This is also associated with entropic penalty $\Delta G_{IR} = T\Delta S_{IR}$ of about 5 kJ/mol per completely restricted rotor at room temperature [15]. Again, most probably the full restriction of bond rotation does not occur and the real entropic penalties are smaller. Another source of energetic penalties that can result in increase of ΔG of binding include structural strains, electrostatic repulsion and hydrophobic groups exposition to the solvent. These terms are difficult to estimate in a general fashion and they have to be considered for each specific case.

The factors that favor binding lead to decrease of ΔG. Such energetic benefits usually come from polar bond formation, various dipole-dipole interactions (also known as Van der Waals forces) and hydrophobic effects. Each individual type of interaction is described in more details in the next section. The polar bond formation is associated with slight increase in entropy, associated with the release of water molecules into bulk water, and strong enthalpic gain, coming from the polar groups interaction. The polar interactions can be formed between ions (ΔG_{ION}) or neutrals (ΔG_{HB}) [16]. Hydrophobic effect is based on the fact that non-polar moieties tend to minimize the surface of contact with the polar solvent [17], which in the case of biomolecules in native environment is water. That makes two hydrophobic groups tend to stick together. The nature of hydrophobic effect is almost completely entropic and can to some extent be related to the area of the of hydrophobic surface buried in the binding interface A_{BHS} [18], so the term ΔG_H can be expressed as = -$A_{BHS}T\Delta S_H$. The Van der Walls forces arise from interactions between induced dipoles, so their nature is enthalpic: $\Delta G_{WdV} = \Delta H_{WdV}$. They are usually the weakest ones and their magnitude decrease very fast with distance [19].

The observations, made in the last two paragraphs, are summarized in the eq. 1.2:

$$\Delta G = \sum \Delta G_{ION} + \sum \Delta G_{HB} + \Delta H_{WdV} - T\Delta S_{T,R} - T\Delta S_{IR} - A_{BHS}T\Delta S_H \quad (1.2)$$

In case there is no positive or negative cooperativity between the elementary interactions, total free energy of binding is a sum of individual contribution of each elementary term. Experimentally derived values for some of these terms, obtained from 45 protein-ligand complexes with known affinities [16], are listed in Table 1.1.

Table 1.1. The values of different contributions to the free energy of binding, derived using 45 protein-ligand complexes with known structures and affinities [16].

Term	**Effect**	**Energy, kJ/mol**
$\Delta G_{T,R}$	Molecular association	+5.4
ΔG_{IR}	Internal rotor restriction	+1.4
$T\Delta S_H$	Hydrophobic effect per Å^2 at room temperature	-0.17
$\Delta G_{HB} + \Delta G_{WdV}$	Polar bonds formed by neutrals	-4.7
ΔG_{ION}	Polar bonds formed by ions	-8.3

1.1.3 Elementary non-covalent interactions

Another way to describe non-covalent interactions between molecules is to break each binding interface into number of elementary interactions, or bonds. To avoid ambiguity, in this text interactions between entire molecules are referred to as “non-covalent interactions”, while elementary bonds that comprise non-covalent interactions will be called “non-covalent bonds”. Non-covalent bond is a pairwise non-covalent interaction between two atoms. One exception is hydrophobic interaction, which is difficult to treat as pairwise, so all the quantitative data will be supplied per unit surface area of non-polar group [20].

There is a number of fundamental differences between covalent and non-covalent bonds. While covalent interactions are caused by the overlap of electronic densities of adjacent atoms and are therefore short-range interactions (the optimal distance is usually about 2 Å, at 4 Å they are practically zero), non-covalent interactions was detected at distances of up to 90 Å [21]. While covalent bonds are formed as a consequence of electron density function overlaps, most of polar non-covalent bonds energy comes from electrostatic, induced dipole-dipole and dispersion interactions [20]. However, research made in the last years suggests that some of the mechanisms that stabilize non-covalent bonds are similar to those found in covalent bonds [22].

The strongest non-covalent bonds, according to the Table 1.1, are formed between two ions. The ionic bond is formed between two atoms whose electronegativity is so different that all the electron cloud is concentrated around “negative” atom and thus cannot be shared with the second atom to form a covalent bond. Instead, they interact electrostatically with the energy, determined by Coulomb's law. The classical example of a system stabilized by ionic interactions is NaCl crystal lattice, and ionic interactions in proteins are often called “salt bridges”, although some authors refer to the mixed ionic/hydrogen bonds as salt bridges [23]. Salt bridges are often found on protein-protein and protein-ligand interactions interfaces [24] and in the protein cores, stabilizing their tertiary structure [25]. In the Table 1.1 ionic interactions are represented by the term ΔG_{ION}.

Another type of polar bonds commonly found in proteins is called hydrogen bonds [26]. Hydrogen bond is formed between hydrogen atom, covalently bound to the electronegative atom and another electronegative atom. The atom covalently bound to hydrogen is called donor, the one that is forming the hydrogen bond – acceptor. The hydrogen bonds are usually 2.6 - 3.1 Å long, and their energy in aqueous environment is estimated to be between 4 and 13 kJ/mol, although these numbers can vary greatly depending on the environment [27],[29].

Classical example of hydrogen bonds is a network, formed by H_2O molecules in liquid water [30]. There are plenty of inter- and intra-molecular bonds found in proteins. Hydroxyl (Ser, Thr, Tyr) or amino (Asn, Gln) groups can form contacts with water, aiding to the solubility of the proteins, or with other groups of the same or other molecule. Peptide bond has both hydrogen bond donor (>NH2) and acceptor (>C=O) groups. Hydrogen bonds between them is a glue that keeps together protein secondary structure motifs, such as helices and β-sheets, and thus is crucial for stability of protein structure [28]. Hydrogen bonds also play key role in intermolecular interactions. It was shown [31], that Asn and Thr, typical hydrogen bond donors, show the highest propensity among all residues to appear on the protein-protein interaction interface. The optimization of hydrogen bond placement is one of the leading strategies for rational drug design [32].

Hydrophobic effect is third major driving force of PPI and PLI. Hydrophobic effect is the tendency of non-polar groups (in biomolecules typically consisting of C and H atoms) to stick together in aqueous media. It is usually explained by the fact that water molecules in the vicinity of hydrophobic group tend to organize in a way that maximizes number of hydrogen bonds they form. Since they cannot form a contact with hydrophobic group, they have to point all the hydrogen donors and acceptors away from its surface. Their motion is thus restricted and the entropy decreased. This creates entropic penalty proportional to the hydrophobic area, exposed to the polar solvent, and hydrophobic groups tend to minimize this area by sticking together [17]. This is most widely accepted explanation, although some authors explain attraction of non-polar groups in aqueous media by van der Waals forces [33]. Hydrophobic patches are very often found on the protein-protein binding interfaces. Hydrophobic surfaces tend to attract to each other independently of their nature, so they are believed to be responsible more for binding affinity than specificity [34].

Van der Waals forces are weak short-range non-specific interactions that occur between atoms of all types when they are put sufficiently close together. They can be approximately described by empirical Lennard-Jones potential [17]. More accurate picture requires quantum calculations [35]. Qualitatively, van der Waals interactions occur either between two induced dipoles or between induced and permanent dipole. The magnitude of van der Waals forces can be up to 4 kJ/mol, but it decreases as distance between dipoles to the power of 6, so they rarely play crucial roles in PPI or PLI, but there are exceptions [36].

1.1.4 Challenges in describing, modulating and designing non-covalent interactions.

The elementary forces, described in the previous section, are the only constituents of the majority of PPI and PLI. If they are known and relatively well-characterized, why cannot we design a ligand to the protein with known three-dimensional structure? Keeping the analogy with the language, one can say that we know the alphabet but we are still unable to communicate. There are at least two major hurdles on the way to better understanding of mechanisms of non-covalent interactions [37], [38]. First, solvating effects, defining how water and dissolved ions affect all the forces that are involved in binding. Second, intrinsic plasticity of large biomolecules, that allows them to deform in ways that are difficult to predict.

Solvent can affect binding in a number of different ways, in addition to hydrophobic interactions, described above. Bulk water has a dielectric constant $\varepsilon = 80$. This is a factor by which energy any long-range electrostatic interaction is decreased in water. When the interacting groups come closer together, the thin layer of water between them cannot be treated as bulk, resulting in lower values of ε. However, even when two ions are placed next to each other with no water molecules between them, the effective dielectric constant of electrostatic interactions is about 4, which is quite different from the $\varepsilon_0 = 1$ (value in vacuum) [39]. This suggests that surrounding water molecules still provide certain level of shielding for the interacting ions. In addition to that, every polar group interacts with the dielectric environment (in this case, solvent). This interaction is characterized by the energy of solvation and results in a substantial energetic penalty for the removal of solvating shell upon binding of the molecule [40]. This is additionally complicated by the fact that the position of other chemical groups also affect the energy of solvation. On top of that, slight variations of pH affect protonation states of interacting molecules that in turn is strongly correlated with the binding affinities [41], [42]. Different models are used to describe these phenomena, such as Poisson-Boltzmann [43] or generalized Born models [44], that work quite successfully for small molecules but often fail to describe interactions that involve conformation flexibility of interacting species.

This brings up second issue that makes it so hard to design efficient binders even if the structure of the target is known. Conformational flexibility is an essential property of large biomolecules, and many protein functions rely on their ability to alter conformation in a response to the external factor (binding of the allosteric ligand, change of pH etc.) [45]. The role of flexibility started to be recognized more than 30 years ago [46], and now there is a strong evidence that conformational dynamics is strongly involved in PPI and PLI [47]. From the perspective of energy balance it means that conformational entropy plays a considerable role in overall energy of binding and has to be accounted for [48], [49]. Several models of interaction between a ligand and a flexible target are considered. In “induced fit” model stable lowest energy protein conformation changes upon binding of the ligand, adapting to its geometry. “Conformer selection” model assume that number of protein conformations with similar energy are in equilibrium, and a ligand selectively binds to one of them, shifting the equilibrium towards the increase of its population. Third model is called “conformational induction” and represents a combination of the first two. The protein samples a number of conformation, and ligand binds to one or several of them, forcing the protein to adapt a conformation it would not have in a non-bound form [50]. Energetically they produce similar results. But geometrically the fact of protein flexibility hampers rational drug design because it offers too many possible protein conformation. The number grows exponentially with a number of degrees of freedom [51] and when flexible ligands are considered [52]. This makes any information about protein flexibility highly valuable for general understanding PPI and PLI and for rational drug design.

Solvation is addressed in the chapters 1-3 of this thesis; the study of protein flexibility is presented in chapter 4.

1.2. Mass-spectrometry

1.2.1 General introduction to mass-spectrometry

Mass-spectrometry (MS) is a technique that allows to measures the ratio between mass and charge of ionic species. It can be applied to a wide range of objects, from isotopes of different elements [53] to small chemicals compounds [54] to larger biological molecules [55], non-covalent complexes [56] and even entire viruses [57]. In all those cases the scheme of the mass-spectrometry experiment remains essentially the same (Figure 1.1).

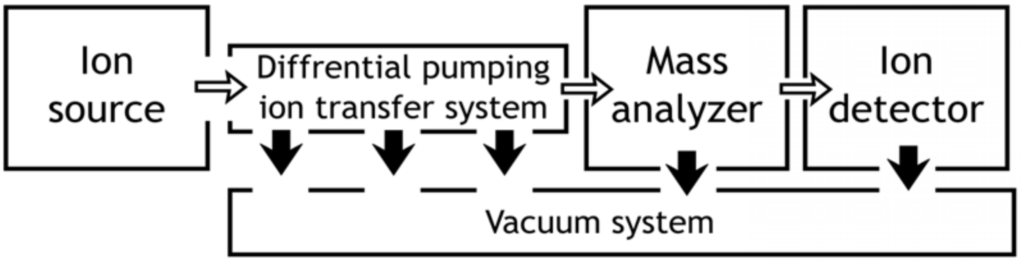


Figure 1.1 The principal scheme of mass-spectrometer.

First, sample in the solid, liquid or gaseous state is vaporized and ionized in the ion source. This usually happens either at atmospheric pressure (atmospheric pressure chemical ionization (APCI), electrospray ionization (ESI) and others) or in low-vacuum (matrix-assisted laser desorption-ionization (MALDI)) [58]. The ions produced in the ion source pass through the series of skimmers into the transfer system, that carries them through different pumping stages until they reach mass analyzer. The ion signal is analyzed and detected either as ion current by the detector (like in time-of-flight (ToF), quadrupole or sector mass-analyzers) or in the form of induced current (Fourier-transform ion cyclotron resonance (FT-ICR), ORBITRAP and the likes). The data is then typically passed to a computer and the output takes form of a mass spectrum – plot of relative intensities of ion signals versus ion mass-to-charge ratio. More about the the state of the art of mass-spectrometry can be found a number of books and reviews, for example [59–62].

1.2.2 Collision-induced dissociation mass-spectrometry.

The vacuum required for mass analyzer can range from 10^{-6} mbar for ToF or quadrupole mass analyzers to 10^{-10} mbar for FT-ICR mass analyzer [63]. In the first instruments the sole purpose of the ion transfer region was to connect low-vacuum inlet region with high-vacuum mass-analyzer section, losing as little ions as possible. But with the advance of technology the transfer region started to play more "active" roles in mass-spectrometry experiment. One popular modification of ion transfer region is ion mobility spectrometry (IMS) cell, that allows to separate ions in the ion beam according to their shape and size prior to mass analysis [64]. IMS cell can provide structural information about the shape of molecules and greatly reduces complexity of the spectra, working as an additional separation step [65]. Various dissociation techniques can be also applied in the transfer region in order to fragment ions and obtain additional structural information or help to identify the analyte [66].

There is a large number of dissociation techniques. They can be based on ion-molecule collision (CID, SORI, HCD), ion-surface collisions (SID), ion irradiation with laser (IRMPD) or black-body radiation (BIRD), electron capture (ECD) or transfer (ETD) and a number of others [67][68]. In spite of the fact that many techniques emerged since the development of CID, it remains the most widely used dissociation method, largely because it is relatively simple and can be applied to a wide range of problems [69], [70].

In a typical CID experiment ions are accelerated by the electric field and introduced in a collision cell that is filled with the inert gas. Noble gases like He, Ar or Xe are often used, as well as N_2 and CO_2. If the ion beam was thermalized before entering the cell, its laboratory frame kinetic energy K_{LAB} is defined by the acceleration voltage V and the charge of the ion z:

$$K_{LAB} = \frac{m_I v_I^2}{2} = zqV \tag{1.3}$$

where q is an elementary charge, v_I and m_I is the ion velocity and mass respectively. The ion undergoes collisions with the target gas molecules and its kinetic energy is transformed into the internal energy E and is deposited in ion's vibrational and rotational degrees of freedom. Maximum energy that can theoretically be deposited in the ion upon single collision is the center of mass kinetic energy of the ion-molecule system K_{COM}:

$$K_{COM} = \frac{\Delta v^2}{2} \frac{m_I m_G}{m_I + m_G} \tag{1.4}$$

where $\Delta v = v_I - v_G$ is a difference in velocities between the gas molecule and the ion and m_G is the mass of the gas molecule [71]. When the internal energy of the ion reaches a certain level, ion starts to dissociate. This energy is referred to as activation energy of unimolecular dissociation E_A [72]. If the energy from a single collision is enough for the fragmentation, the method is dubbed "single collision CID", otherwise it is called "multiple collision CID" [73]. In case of large difference between the masses of the ion and the gas molecule K_{COM} can be much smaller than K_{LAB}. Furthermore, the internal energy needed for the same amount of energy per degree of freedom increases with the size of the molecule. In case of experiments on non-covalent complexes one usually wants to preserve covalent bonds intact. It limits the maximum K_{COM} that can be achieved in a single collision. Thus the majority of CID experiments involving non-covalent complexes are performed in a multiple collision regime [74].

Although the upper limit for the energy deposited in the ion upon collision is well defined, the actual magnitude is still under debate. The ion excitation depends on COM kinetic energy of the system and is characterized by the energy transfer efficiency η - the average fraction of K_{COM} transformed into internal energy of the ion [75]:

$$\Delta E = \eta K_{COM} \tag{1.5}$$

where ΔE is a positive change of ion internal energy, or energy step. η in turn is a function of kinetic energy with the average number calculated for small peptide being 12.8 ± 2.1% [76] [77]. Since it is the most accurate value reported to-date, it is suggested to use it for the treatment of other peptides and proteins [75]. Apart from that, energy deposition in the molecular ion depends on its current internal energy. This dependence is governed by the density of energy states of the ion, the function that is extremely difficult to calculate for proteins of considerable size (10s of kDa) [78]. On top of that, collision with the gas molecule can result in the decrease of the internal energy of an ion – so called collisional cooling [79],[80].

A number of existing techniques can successfully describe collisional activation of small molecules [81],[82] or small peptides [77] prior to fragmentation. However, for the larger ions like proteins or protein complexes there is no general solution for this problem.

1.2.3 Native mass-spectrometry

Native mass-spectrometry is a new fast emerging technique that allows the investigation of proteins and entire protein complexes in gas phase while keeping their state close to native [83].

The native MS is quickly gaining pace as a tool for biochemistry [84], structural biology [85] and drug discovery [86]. The method takes advantage of “soft” ionization techniques, that are capable of ionization and evaporation of biological molecules with minimal disturbance of their native structure. Such techniques include MALDI [87] and, more commonly, ESI and its derivatives (nano-ESI [88], or electrosonic spray ionization ESSI [89]). Nano-ESI is especially popular thanks to its sub-µl/min sample consumption rates. Although native MS does not provide structural information at atomic resolution, its speed, sensitivity and ability to detect several species in one mixture makes it a powerful instrument for investigation of multiprotein complex organization and interaction pathways [90], [91].

Native mass-spectrometry of protein-ligand complexes opens up possibilities for high-throughput screening methodologies [86] as well as for the study of individual non-covalent complexes in gas-phase [92–95]. The gas-phase studies provide a valuable information about the role that the solvent plays in non-covalent binding. Gas-phase stability of the complex can be compared with solution binding energies in order to determine the impact of water screening, solvation and/or hydrophobic interactions. However, these studies are impeded by the inability to accurately quantify the internal energy that is needed to dissociate non-covalent complex in gas phase. Solution to this problem could open the way towards quantitative description of gas phase non-covalent chemistry.

1.2.4 Non-covalent bonds and non-covalent interactions in gas phase

Today there is a strong evidence that proteins, ionized by ESI, largely preserve their native conformations in gas phase, at least on a time scale of typical MS experiment (<10 ms) [96], [97], although native structure in gas phase does not represent a global minimum of energy [98]. The ions, produced by ESI, are formed by the gradual evaporation of solvent from nanometers-sized droplets, each containing single protein or protein complex. Eventually charged protein remains with no or little water molecules [99]. The water molecules, that are coordinated around charged atoms, are the last to dissociate, thus helping to maintain native structure intact in gas phase [100]. The evaporation of last water molecules is followed by the collapse of the structure thanks to instantaneous increase in number and strength of electrostatic interactions, caused by the removal of screening solvation shell [101]. The resulting structure retains native-like features, that is supported by the evidence that proteins remain fully functional after undergoing soft-landing in mass-spectrometer [102]. It was shown that α-helices are stabilized by the collapsed charged side-chains thanks to their interactions with dipoles of hydrogen bonds [103]. In addition, the strength of the hydrogen bond itself is expected to become stronger – close to the that under isolated conditions, that were shown to be ≈5-6 kJ/mol [29].

On a time scale from 10 of ms to seconds irreversible processes are starting to occur, the main reason being the fact that hydrophobic interactions are dramatically reduced in gas phase

[104], [105]. It was shown on for the cytochrome *c* that the regions that are stabilized by hydrophobic bonds and are the core stable regions in solution become the least stable in the gas phase [106]. As a result, the thermodynamic equilibrium is shifted by the weakening of hydrophobic interactions and strengthening of electrostatics, and the structure starts to unfold in order to find new most stable structure. Which may be achieved in the course of the next 10 – 100 seconds [107]. This time is far beyond the time scale of most MS experiments. In conclusion, proteins are likely not to lose their native-like structure before they are detected in a typical MS experiment.

The implications that it has on PPI and PLI in gas phase can be summarized as follows. Although non-covalent complexes are relatively fragile structures, and “there is a very fine line between sufficient desolvation of the gas-phase complex and dissociation of the complex.” (J, Loo, [108]), there is a wealth of experimental evidence that suggests that non-covalent complexes are preserved in the gas-phase. One reason can be that new electrostatic interaction that are formed in gas phase can to some extent stabilize the complex. Another possibility is that if hydrophobic interactions, as suggested in [33], indeed are partly caused by van der Waals forces, they could retain some attractive force even after removal of the solvent. The accurate description of this phenomena is still under debate. One fact that is clear today: the gas phase non-covalent binding energy cannot be taken as a quantitative estimate of solution binding energy [109], [110]. That opens up a possibility to investigate the role of the solvent in non-covalent interactions by comparison of the solution and gas-phase binding energies. In order to do a quantitative analysis, a good method for calculation of gas-phase binding energy is needed. This question is addressed in chapter two of the present thesis.

1.3 Role of dynamics in PPI and PLI

1.3.1 Biological role

It is a well established fact that large biological molecules, in particular proteins, are highly dynamic systems [111]. It has immediate implications for protein-protein and protein-ligand recognition processes [112], [113]. As it was discussed in section 1.1.4, the change of paradigm from simple static “lock-and-key” model to the “induced fit model” and further to “conformational selection” model provided better understanding of binding processes but also brought numerous complications into the tasks of docking and rational ligand design, because protein and ligand plasticity is extremely difficult to account for [114]. Things look even worse if one wants to describe interactions involving intrinsically disordered proteins (IDP) that are now believed to constitute about 30% of genome-encoded proteins in eukaryotes [115]. A number of computational approaches are being actively developed [116–118], but experimental data still remains the only source of reliable information about the dynamics of proteins, that is used to parametrize and validate computational models.

1.3.2 NMR as a tool to study dynamics of biomolecules.

1.3.2.1 Basic theory

NMR relaxation experiments are used to characterize molecular motion at time scales ranging from picoseconds to seconds [119]. The excited nuclear spin can undergo relaxation through two distinct processes. The first relaxation process, called longitudinal relaxation, is associated with returning of magnetic moment into relaxed state, i.e. re-aligning with the constant magnetic field of the spectrometer. During this process the energy is transformed from the excited states to the surrounding environment of the nuclei, or “lattice”, so this mechanism is also called “spin-lattice relaxation”.

The second relaxation process is called transverse relaxation. It involves loss of the coherence between excited nuclear spins, as they are precessing around static magnetic field. This process arises mainly from dipolar coupling between magnetic dipoles and chemical shift anisotropy (inhomogeneity in density of electrons surrounding the nuclei). Since the first process plays a major role here, the process is often referred to as spin-spin relaxation. This process does not involve energy exchange with lattice, only between the magnetic dipoles.

Both processes are usually described as a first-order exponential decay (eq. 1.6):

$$I(t) = e^{-\frac{t}{T_{1,2}}} \tag{1.6}$$

where *I(t)* is a signal intensity as a function of time *t*, and T_1 and T_2 are spin-lattice and spin-spin relaxation times respectively. For large molecules like proteins values of T_1 are usually around 1 s or higher, T_2 is in the range of tens to hundreds of ms.

Another experimentally measured quantity that is used in protein dynamics relaxation studies is heteronuclear Overhauser effect (hnNOE). The Nuclear Overhauser Effect (NOE) indicates how the populations of ground and excited states of one spin are influenced by populations of the other spin, coupled to it. In the case of heteronuclear coupling, NOE effect measures the populations of the spin coupled to proton when that proton is saturated. Usually the ratio of the signal intensity with and without proton saturation is taken as a measure of NOE.

Nuclei that are used most often for relaxation studies of proteins are ^{15}N and ^{13}C, because in their case relaxation processes are dominated by CSA and dipolar coupling to the bonded hydrogen atom. The relaxation effects caused by these interactions can be calculated with good precision, and the analysis of experimental data is more accurate.

The basic equations, correlating spectral densities with relaxation times for the case of ^{1}H - ^{15}N dipolar interactions are the following:

$$R_1 = \frac{1}{T_1} = 3\left(d^2 + c^2\right) J\left(\omega_N\right) + d^2\left[3J\left(\omega_H - \omega_N\right) + 6J\left(\omega_H + \omega_N\right)\right] \tag{1.7}$$

$$R_2 = \frac{1}{T_2} = \frac{1}{2}\left(d^2 + c^2\right)\left[4J\left(\omega_0\right) + 3J\left(\omega_N\right)\right] + \frac{1}{2}d^2\left[J\left(\omega_H - \omega_N\right) + 6J\left(\omega_H\right) + 6J\left(\omega_H + \omega_N\right)\right] + R_{ex}$$

(1.8)

$$hnNOE = 1 - \left|\gamma_H/\gamma_N\right| d^2\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right]/R_1 \qquad (1.9)$$

where R_1 and R_2 are longitudinal and transverse relaxation rates respectively, $J(\omega)$ is a spectral density at frequency ω; ω_H, ω_N, γ_H and γ_N are Larmor frequencies and gyromagnetic ratios of ^1H and ^{15}N nuclei, respectively, and R_{ex} is a conformational exchange rate constant, that contributes to R_2. Parameters d and c are calculated according to equations 1.10 and 1.11 and come from contributions of dipolar coupling between ^1H and ^{15}N nuclei and ^{15}N CSA, respectively.

$$d = -\left[\frac{\mu_0}{4\pi}\right]\gamma_H\gamma_N\frac{h}{4\pi r_{NH}^3} \qquad (1.10)$$

$$c = \frac{-\omega_N\Delta\sigma}{3} \qquad (1.11)$$

where μ_0 is a magnetic constant, h is Planck constant, r_{NH} is the length of NH bond and $\Delta\sigma$ is a CSA of ^{15}N nucleus.

The experimental data treatment used for protein backbone dynamics determination with NMR generally involves the following steps:

1. R_1, R_2 and hnNOE of the protein marked with ^{15}N are determined experimentally for each backbone amide that is represented by a signal on the HSQC spectrum.
2. The spectral densities $J(\omega)$ are calculated for set of frequencies using equations 1.7 – 1.11 and various assumptions, described below [120], [121].
3. The calculated spectral densities are used to deduce the picture of protein dynamics.

1.3.2.2 Reduced spectral density mapping

One rather straightforward way to extract information about protein backbone dynamics from spectral densities applicable to ^{15}N relaxation dynamics is called reduced spectral density mapping (RSDM) [122]. The spectral densities $J(\omega)$ can be calculated at three points using equations 1.7 - 1.11 and set of assumptions, suggested in [123]:

1. $J(\omega)$ is a slowly varying function of ω between $(\omega_H - \omega_N)$ and $(\omega_H + \omega_N)$
2. $J(\omega) \propto 1/\omega^2$ when $\omega \approx \omega_H$;

Then if there are $J(\beta_1\omega_H)$ and $J(\beta_2\omega_H)$, then they can be related to each other as follows:

$$J(\beta_1\omega) = \left(\frac{\beta_2}{\beta_1}\right)^2 J(\beta_2\omega) \qquad (1.12)$$

where $\beta_1, \beta_2 \approx 1$. Since ω_N is about ten times smaller than ω_H, the latter holds true for ω if $(\omega_H - \omega_N) \lessapprox \omega \lessapprox (\omega_H + \omega_N)$. It is shown in [123] that equations 1.7-1.9 can be rearranged and simplified by approximating three terms $J(\omega_H)$, $J(\omega_H - \omega_N)$ and $J(\omega_H + \omega_N)$ with terms of the form $\alpha J(\beta\omega_H)$ and assuming that $R_{ex} = 0$:

$$J(0.87\omega_H) = \frac{1}{5d^2}\left|\frac{\gamma_N}{\gamma_H}\right|(1 - NOE)R_1 \tag{1.13}$$

$$J(\omega_N) = \frac{R_1 - 7d^2 J(0.921\omega_H)}{3(d^2 + c^2)} \tag{1.14}$$

$$J(0) = \frac{2R_2 - R_1 - 6d^2 J(\omega_H)}{4(d^2 + c^2)} \tag{1.15}$$

Applying equation 1.12, further simplification can be obtained: $J(\omega_H) = (0.87)2J(0.87\omega_H) = 0.757J(0.87\omega_H)$ and $J(0.921\omega_H) = (0.87/0.921)2J(0.87\omega_H) = 0.892J(0.87\omega_H)$. Combining the latter with equations 1.14 and 1.15, the final set of equations can be obtained:

$$J(0.87\omega_H) = \frac{1}{5d^2}\left|\frac{\gamma_N}{\gamma_H}\right|(1 - NOE)R_1 \tag{1.16}$$

$$J(\omega_N) = \frac{R_1 - 6.244d^2 J(0.87\omega_H)}{3(d^2 + c^2)} \tag{1.17}$$

$$J(0) = \frac{2R_2 - R_1 - 4.542d^2 J(0.87\omega_H)}{4(d^2 + c^2)} \tag{1.18}$$

Equations 1.16 – 1.18 allow determination of the spectral density at three frequencies ($\omega = 0$, ω_N and $0.87\omega_H$) for each backbone ^{15}N nucleus. Then $J(0)$ can be attributed to the “slow” motion of the protein backbone, while $J(0.87\omega_H)$ and $J(\omega_N)$ - to the “fast” motion on the ns time scale. For 600 MHz spectrometer $\omega_N \approx$ 60.7 MHz and $0.87\omega_H$ = 522 MHz, so the spectral density at these frequencies provide insight into molecular motion at 16.5 ns and 1.9 ns respectively.

The RSDM is an attractive approach because it is very straightforward to apply. However, it has a number of intrinsic drawbacks, that complicate direct interpretation of the results. One problem lies in the assumption that the value of CSA is constant across the protein. Values that are usually taken for the analysis are between -172 and -160 ppm, that come from solid-state NMR. However, those values are can be widely distributed. For example, for ubiquitin it was shown that CSA can spread from -216 to -125 ppm [124], [125]. This renders inaccurate the information about protein dynamics obtained from spectral density mapping with fixed value of CSA.

This issue can be addressed with the multi-field approach, when relaxation data is recorded at several different magnetic field strengths [126]. The site-specific values of CSA can then be obtained thanks to the fact that $\Delta\sigma$ and $J(0)$ are independent on magnetic field strength. Considering that R_{ex} is known or 0, site-specific $\Delta\sigma$ can be extracted if the experimental relaxation data at only two different field strengths is available, because equations 1.16-1.18, constructed for two datasets, will result in six equations with six unknowns.

Another problem is the assumption that chemical exchange does not affect the experimentally determined relaxation rates, i.e. $R_{ex} = 0$. Chemical exchange is a term, used in NMR studies to refer to the processes that involve molecular motion on a μs to ms time scale [127], like interdomain motion, loop motion etc. [128],[125]. In the RSDM framework chemical exchange can contribute to *J(0)*, calculated from eq. 1.18 that in turn was derived from eq. 1.8 [119]. That can lead to overestimation of *J(0)* and, more importantly, to loss of important biological information that is usually associated with such kind of motion. RSDM approach allows to partially overcome this problem by, again, utilizing data obtained at different magnetic field strengths [126]. The suggested strategy relies on the assumption that $R_{ex} \propto B_0^2$. The *J(0)* is calculated at two field strengths, function of *J(0)* vs B_0^2 is estimated and "real" *J(0)* is calculated at the point where $B_0^2 = 0$. This approach eliminates errors in the *J(0)* calculation, but it fails to predict the R_{ex}. More complex approach, that involves assumptions about the distributions of spectral densities, allows direct determination of R_{ex} for the rigid part of the protein [124]. But the most interesting processes, that involve chemical exchange, usually tend to occur in more flexible parts of proteins, making this approach less useful.

1.3.2.3 Model-free formalism

Further source of errors arising while trying to analyze NMR relaxation data with RSDM is the fact that the motion of each nucleus is a combination of fast local motion and slow overall rotation and translation molecule, often called "tumbling". Since the flexibility of protein molecule is governed by the degree of local motion, allowed for its groups, it is desirable to separate it from the molecular tumbling. Model-free formalism [129], [130] is attempting to do just that.

In model-free (MF) formalism it is assumed that if rotational diffusion tensor of the molecule is known, the contribution of molecular tumbling can be subtracted from the spectral density, leaving only the part that arises from internal motion of the molecule. In classical approach, the internal motion is then characterised by two parameters, that does not depend on the type of motion: order parameter S and correlation time of motion τ_e. S (usually treated in the form of S^2) is a measure of relative amplitude of local motion that scales from 0 to 1, where $S^2 = 0$ corresponds to completely unrestricted motion and $S^2 = 1$ is expected in when there is no free movement at all. More widely used nowadays is extended model-free formalism (EMF), suggested in [131]. In EMF more complex internal dynamics picture is utilized, decomposing local motion into "fast" and "slow" component. Four parameters are used to characterise internal motion: "fast" and "slow" correlation times τ_f *and* τ_s, $\tau_f \ll \tau_s$, and two order parameters S_s and S_f, $S^2 = S_s^2 S_f^2$. Both τ_s and τ_f are much smaller than the overall rotational correlation time τ_c. In EMF spectral density is expressed through these parameters as follows:

$$J(\omega) = \frac{S^2\tau_c}{1+(\omega\tau_c)^2} + \frac{1-S_f^2)\tau_f}{1+(\omega\tau_f)^2} + \frac{(S_f^2 - S^2)\tau_s}{1+(\omega\tau_s)^2} \qquad (1.19)$$

In addition to four parameters mentioned, R_{ex} can also be considered in EMF formalism to account for a contribution of chemical exchange in the measured values of R_2 [132].

Before the EMF formalism can be successfully applied the overall rotational correlation time τ_c has to be calculated. In the simplest case of isotropic rotation of the molecule the correlation time is calculated as $\tau_c = 1/6D_{iso}$, where D_{iso} is the diffusion coefficient of random Brownian rotation of the molecule. In many real cases shape of a molecule is far from spherical and more sophisticated models for its rotational motion are needed [133]. In a general case of fully anisotropic motion diffusion coefficient is expressed as 3x3 tensor, called rotational diffusion tensor. It can be diagonalised, resulting in three diagonal elements D_x, D_y and D_z. If the shape of the protein can be approximated with axially symmetric model, the number of independent tensor elements is further reduced to 2: $D_\perp$ and $D_\|$.

In order to fully characterize rotational diffusion of the molecule, six parameters have to be found: three diagonal components of rotational diffusion tensor and three angles, needed to align the molecule so that the off-diagonal components of the tensor are equal to zero. The approach that is mostly used for that relies on R_2/R_1 ratios, calculated for the nuclei that are situated in a rigid part of macromolecule [134]. This approach requires previous knowledge of 3-dimensional structure of the molecule, ideally – high resolution X-ray or NMR structure. The R_2/R_1 ratios are estimated from some initial parameters and the systematic search is conducted in the six-dimensional space to find the values that provide the smallest discrepancy between between observed and generated values. Other approaches are suggested to speed up the analysis, for example [135].

When overall rotational motion of the molecule is fully characterized, the EMF approach can be applied. Series of models is constructed, each utilizing different set of model-free parameters. The maximum number of parameters allowed for the model is determined by the number of experiments conducted at different magnetic field strengths. For example, if R_1, R_2 and NOE were determined at one static magnetic field, no more than 3 parameters can be used in each model. In classical approach, implemented by A. Palmer in widely used ModelFree 4.2 program, five possible sets of model-free parameters are considered: model 1 (S_s^2), model 2 (S_s^2 and τ_e), model 3 (S_s^2 and R_{ex}), model 4: (S_s^2, τ_e and R_{ex}) and model 5 (S_f^2, S_s^2, and τ_e). The spectral density function of each spin fitted in each of five models that produces less error is picked as the most adequate one. Similar approach is utilized by the different software, made for model-free analysis of NMR relaxation data. To name a few, TENSOR 2.0 program by P. Dosset at. al [136] allows determination of model-free parameters from relaxation data acquired at one of two different static magnetic field strengths. Unlike ModelFree 4.2 where only isotropic and axially symmetric rotational diffusion tensor is considered, TENSOR 2.0 allows to chose between fully isotropic, axially symmetric and fully anisotropic rotational diffusion tensors on the basis of graphical representation of fitting results. DYNAMICS 3.0 is another implementation of similar algorithm, produced by D. Fushman et. al. [137–139]. Similar to TENSOR 2.0, it accepts fully anisotropic diffusion tensor and multiple static magnetic fields. In addition to that, thanks to advanced model-selection algorithms, it is notably faster than previous two algorithms. It is implemented as a set of Matlab scripts, that makes the code transparent and easy to modify.

1.4. Vascular endothelial growth factor

VEGF family has seven members: VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGF-E, VEGF-F, and placental growth factor PGF. Members of VEGF family share many structural features that can be described on the example of VEGF-A [140]. VEGF-A is a globular covalent antiparallel homodimer, kept together by 2 intermolecular disulfide bridges and hydrophobic attraction. Each monomer has at least 6 intramolecular disulfide bridges, that form highly stable cystine knot motif [141]. For the work presented in this thesis we used 11-109 construct of VEGF-A121 (Figure 1.2) . It was used because of its higher solubility in comparison with the full length protein and the availability of a plasmid. In spite of being truncated, this construct of VEGF does not show any decrease of activity [142]. It is a highly structured protein, with secondary structure content of about 12% helical and 53% β-sheets. In addition to that each monomer features two loop regions, D63-G65 (loop 1) and P85-G88 (loop 2). Residues of both loop regions and N-terminal α-helix (F17-Y25) are involved in interactions between VEGF and its receptors [143–145].

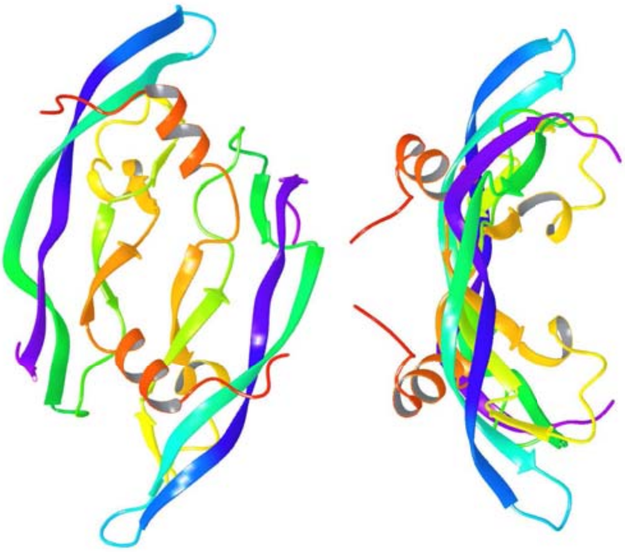


Figure 1.2. X-ray structure of 11-109 construct of VEGF-A, “top” and “side” view.

Vascular endothelial growth factor (VEGF) is a key proangiogenic agent, mainly responsible for the growth of new and re-growth of damaged vasculature [146], although recent reports suggest it also has “non-vascular” roles [147]. In humans VEGF-A is expressed in six forms that differ in number of amino-acids: 121, 145, 165, 183, 189 and 206. All forms of VEGF-A interact with two tyrosine-kinase receptors, VEGFR-1 (Flt-1) and VEGFR-2 (KDR/Flk-1) which are found mainly in endothelial cells. In addition to that, 165 amino-acid form - VEGF-A_{165} – can interact with neuropilin-1 and neuropilin-2 coreceptors [148].

The up-regulation of VEGF-A is associated with tumor growth. Tumoral cells express large amounts of VEGF inducing blood vessel spread in the tumor and facilitating blood supply to it. Inhibition of VEGF-induced angiogenesis is a recognized therapeutic strategy, that was shown to suppress tumor proliferation [149] and increase its response to radiotherapy [150]. This makes VEGF a potent therapeutic target and encourages better understanding of VEGF-receptor interactions and development of new inhibitors [151].

1.5 GroEL

GroEL is a protein, that is found in large number of bacteria. It belongs to the chaperonin family of molecular chaperones. Its main function is to assist protein folding in the cell and repair misfolded proteins [152]. *e. coli* GroEL together with its helper-protein GroES are the most studied chaperonins [153]. GroEL consists of two 7-member rings, stacked on top of each other, forming a barrel-like structure. The mass of each of 14 subunits is around 57 kDa, so the total mass of the complex is around 800 kDa. Mode of function of GroEL can be generally described in the following way [154]. The unfolded (or misfolded) substrate is encapsulated in the central cavity of GroEL, which is then capped by the GroES heptamer. Inside the cavity folding (or refolding) of the substrate occurs. The exact mechanism of this process is not known. The encapsulation and release of substrate is associated with conformational change of GroEL and is mediated by the allosteric binding and hydrolysis of ATP.

Another aspect of the mechanism of function of GroEL is the mechanism of allosteric interaction between GroEL and ATP. Traditionally two allosteric regulation models are considered: (1) the Monod-Wyman-Changeux (MWC) model, conformational changes take place in a concerted manner [155]. Applyed to GroEL that means that in MWC model all subunits within one ring undergo conformational change simultaneously. (2) The Koshland-Némethy-Filmer (KNF) model in which conformational changes occur sequentially, that is, subunits within one ring change their conformation one by one [156]. The ability to distinguish between two mechanisms could shed light on the structural mechanism and thermodynamics of chaperone-assisted protein folding in prokaryotes and eukaryotes [157].

In the context of the facts presented above in the present thesis we have addressed the following three objectives:

1. ***To develop a combined experimental/computational approach for quantitative assessment of gas-phase stabilities of non-covalent complexes.***
2. ***To develop methodology for study of allosteric regulation of large protein complexes using structural mass-spectrometry***
3. ***To study backbone dynamics of VEGF using combination of NMR relaxation and normal mode analysis***

References:

[1] M. Baker, “Proteomics: The interaction map,” *Nature*, pp. 2–5, 2012.

[2] S. Gordo and E. Giralt, “Knitting and untying the protein network: modulation of protein ensembles as a therapeutic strategy.,” *Protein science : a publication of the Protein Society*, vol. 18, no. 3, pp. 481–93, Mar. 2009.

[3] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E. E. Wanker, “A human protein-protein interaction network: a resource for annotating the proteome.,” *Cell*, vol. 122, no. 6, pp. 957–68, Sep. 2005.

[4] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal, “Towards a proteome-scale map of the human protein-protein interaction network.,” *Nature*, vol. 437, no. 7062, pp. 1173–8, Oct. 2005.

[5] Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter, T. Maniatis, A. Califano, and B. Honig, “Structure-based prediction of protein-protein interactions on a genome-wide scale.,” *Nature*, vol. 490, no. 7421, pp. 556–60, Oct. 2012.

[6] M. P. H. Stumpf, T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe, and C. Wiuf, “Estimating the size of the human interactome.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 19, pp. 6959–64, May 2008.

[7] L. A. N. Amaral, “A truer measure of our ignorance.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 19, pp. 6795–6, May 2008.

[8] S. B. Zimmerman and A. P. Minton, “Macromolecular crowding: biochemical, biophysical, and physiological consequences.,” *Annual review of biophysics and biomolecular structure*, vol. 22, pp. 27–65, Jan. 1993.

[9] P. L. Kastritis, I. H. Moal, H. Hwang, Z. Weng, P. A. Bates, A. M. J. J. Bonvin, and J. Janin, “A structure-based benchmark for protein-protein binding affinity.,” *Protein science : a publication of the Protein Society*, vol. 20, no. 3, pp. 482–91, Mar. 2011.

[10] G. Klebe and H. J. Böhm, “Energetic and entropic factors determining binding affinity in protein-ligand complexes.,” *Journal of receptor and signal transduction research*, vol. 17, no. 1–3, pp. 459–73.

[11] D. H. Williams, E. Stephens, D. P. O’Brien, and M. Zhou, “Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes.,” *Angewandte Chemie (International ed. in English)*, vol. 43, no. 48, pp. 6596–616, Dec. 2004.

[12] K. Sharp, “Entropy-enthalpy compensation: fact or artifact?,” *Protein science : a publication of the Protein Society*, vol. 10, no. 3, pp. 661–7, Mar. 2001.

[13] S. H. Gellman, “Introduction: Molecular Recognition.,” *Chemical reviews*, vol. 97, no. 5, pp. 1231–1232, Aug. 1997.

[14] W. P. Jencks, “On the attribution and additivity of binding energies,” *Proceedings of the National Academy of Sciences*, vol. 78, no. 7, pp. 4046–4050, Jul. 1981.

[15] M. I. Page, “Entropy, Binding Energy, and Enzymic Catalysis,” *Angewandte Chemie International Edition in English*, vol. 16, no. 7, pp. 449–459, Jul. 1977.

[16] H.-J. Böhm, “The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure,” *Journal of Computer-Aided Molecular Design*, vol. 8, no. 3, pp. 243–256, Jun. 1994.

[17] A. V. Finkelstein and O. Ptitsyn, *Protein Physics: A Course of Lectures (Soft Condensed Matter, Complex Fluids and Biomaterials)*. Academic Press, 2002, p. 354.

[18] N. Muller, “Search for a realistic view of hydrophobic effects,” *Accounts of Chemical Research*, vol. 23, no. 1, pp. 23–28, Jan. 1990.

[19] C. M. Roth, B. L. Neal, and A. M. Lenhoff, “Van der Waals interactions involving proteins.,” *Biophysical journal*, vol. 70, no. 2, pp. 977–87, Feb. 1996.

[20] P. Hobza and K. Mller-Dethlefs, *Non-Covalent Interactions: Theory and Experiment (RSC Theoretical and Computational Chemistry Series)*. Royal Society of Chemistry, 2009, p. 238.

[21] A. Mittal, B. Jayaram, S. Shenoy, and T. S. Bawa, “A stoichiometry driven universal spatial organization of backbones of folded proteins: are there Chargaff’s rules for protein folding?,” *Journal of biomolecular structure & dynamics*, vol. 28, no. 2, pp. 133–42, Oct. 2010.

[22] E. Isaacs, A. Shukla, P. Platzman, D. Hamann, B. Barbiellini, and C. Tulk, “Covalency of the Hydrogen Bond in Ice: A Direct X-Ray Measurement,” *Physical Review Letters*, vol. 82, no. 3, pp. 600–603, Jan. 1999.

[23] E. V. Anslyn and D. A. Dougherty, *Modern Physical Organic Chemistry*. University Science, 2005, p. 1104.

[24] D. Xu, C. J. Tsai, and R. Nussinov, “Hydrogen bonds and salt bridges across protein-protein interfaces.,” *Protein engineering*, vol. 10, no. 9, pp. 999–1012, Sep. 1997.

[25] S. Kumar and R. Nussinov, “Salt bridge stability in monomeric proteins.,” *Journal of molecular biology*, vol. 293, no. 5, pp. 1241–55, Nov. 1999.

[26] D. Xu, C. J. Tsai, and R. Nussinov, “Hydrogen bonds and salt bridges across protein-protein interfaces,” *Protein Engineering Design and Selection*, vol. 10, no. 9, pp. 999–1012, Sep. 1997.

[27] M. Huš and T. Urbic, “Strength of hydrogen bonds of water depends on local environment.,” *The Journal of chemical physics*, vol. 136, no. 14, p. 144305, Apr. 2012.

[28] C. N. Pace, “Energetics of protein hydrogen bonds.,” *Nature structural & molecular biology*, vol. 16, no. 7, pp. 681–2, Jul. 2009.

[29] S.-Y. Sheu, D.-Y. Yang, H. L. Selzle, and E. W. Schlag, “Energetics of hydrogen bonds in peptides.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 22, pp. 12683–7, Oct. 2003.

[30] S. J. Suresh and V. M. Naik, “Hydrogen bond thermodynamic properties of water from dielectric constant data,” *The Journal of Chemical Physics*, vol. 113, no. 21, p. 9727, Dec. 2000.

[31] A. Fernández, L. R. Scott, and H. A. Scheraga, “Amino Acid Residues at Protein−Protein Interfaces: Why Is Propensity so Different from Relative Abundance?,” *The Journal of Physical Chemistry B*, vol. 107, no. 36, pp. 9929–9932, Sep. 2003.

[32] R. Patil, S. Das, A. Stanley, L. Yadav, A. Sudhakar, and A. K. Varma, “Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing.,” *PloS one*, vol. 5, no. 8, p. e12029, Jan. 2010.

[33] C. J. Oss, D. R. Absolom, and A. W. Neumann, “The ‘hydrophobic effect’: Essentially a van der Waals interaction,” *Colloid and Polymer Science Kolloid Zeitschrift & Zeitschrift fur Polymere*, vol. 258, no. 4, pp. 424–427, Apr. 1980.

[34] D. R. Davies, A. D. Gelinas, C. Zhang, J. C. Rohloff, J. D. Carter, D. O’Connell, S. M. Waugh, S. K. Wolk, W. S. Mayfield, A. B. Burgin, T. E. Edwards, L. J. Stewart, L. Gold, N. Janjic, and

T. C. Jarvis, “Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 49, pp. 19971–6, Dec. 2012.

[35] J. Tao, J. P. Perdew, and A. Ruzsinszky, “Accurate van der Waals coefficients from density functional theory.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 1, pp. 18–21, Jan. 2012.

[36] D. Leckband, “Measuring the forces that control protein interactions.,” *Annual review of biophysics and biomolecular structure*, vol. 29, pp. 1–26, Jan. 2000.

[37] M. K. Gilson and H.-X. Zhou, “Calculation of protein-ligand binding affinities.,” *Annual review of biophysics and biomolecular structure*, vol. 36, pp. 21–42, Jan. 2007.

[38] G. M. Whitesides and V. M. Krishnamurthy, “Designing ligands to bind proteins,” *Quarterly Reviews of Biophysics*, vol. 38, no. 04, pp. 385–395, 2005.

[39] G. Löffler, H. Schreiber, and O. Steinhauser, “Calculation of the dielectric properties of a protein and its solvent: theory and a case study.,” *Journal of molecular biology*, vol. 270, no. 3, pp. 520–34, Jul. 1997.

[40] M. K. Gilson and B. Honig, “Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis.,” *Proteins*, vol. 4, no. 1, pp. 7–18, Jan. 1988.

[41] M.-S. Park, C. Gao, and H. A. Stern, “Estimating binding affinities by docking/scoring methods using variable protonation states.,” *Proteins*, vol. 79, no. 1, pp. 304–14, Jan. 2011.

[42] W. W. Kwok, H. Reijonen, B. A. Falk, D. M. Koelle, and G. T. Nepom, “Peptide binding affinity and pH variation establish functional thresholds for activation of HLA-DQ-restricted T cell recognition.,” *Human immunology*, vol. 60, no. 7, pp. 619–26, Jul. 1999.

[43] A. Sayyed-Ahmad, K. Tuncay, and P. J. Ortoleva, “Efficient solution technique for solving the Poisson-Boltzmann equation.,” *Journal of computational chemistry*, vol. 25, no. 8, pp. 1068–74, Jun. 2004.

[44] W. Im, M. S. Lee, and C. L. Brooks, “Generalized born model with a simple smoothing function.,” *Journal of computational chemistry*, vol. 24, no. 14, pp. 1691–702, Nov. 2003.

[45] K. Henzler-Wildman and D. Kern, “Dynamic personalities of proteins.,” *Nature*, vol. 450, no. 7172, pp. 964–72, Dec. 2007.

[46] G. Weber, *Advances in Protein Chemistry Volume 29*, vol. 29, no. null. Elsevier, 1975.

[47] A. V Smrcka, N. Kichik, T. Tarragó, M. Burroughs, M.-S. Park, N. K. Itoga, H. A. Stern, B. M. Willardson, and E. Giralt, “NMR analysis of G-protein betagamma subunit complexes reveals a dynamic G(alpha)-Gbetagamma subunit interface and multiple protein recognition modes.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 2, pp. 639–44, Jan. 2010.

[48] C. Diehl, O. Engström, T. Delaine, M. Håkansson, S. Genheden, K. Modig, H. Leffler, U. Ryde, U. J. Nilsson, and M. Akke, “Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3.,” *Journal of the American Chemical Society*, vol. 132, no. 41, pp. 14577–89, Oct. 2010.

[49] T. S. G. Olsson, M. A. Williams, W. R. Pitt, and J. E. Ladbury, “The thermodynamics of protein-ligand interaction and solvation: insights for ligand design.,” *Journal of molecular biology*, vol. 384, no. 4, pp. 1002–17, Dec. 2008.

[50] S. J. Teague, “Implications of protein flexibility for drug discovery.,” *Nature reviews. Drug discovery*, vol. 2, no. 7, pp. 527–41, Jul. 2003.

[51] I. L. Alberts, N. P. Todorov, and P. M. Dean, “Receptor flexibility in de novo ligand design and docking.,” *Journal of medicinal chemistry*, vol. 48, no. 21, pp. 6585–96, Oct. 2005.

[52] M. P. Repasky, M. Shelley, and R. A. Friesner, “Flexible ligand docking with Glide.,” *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, vol. Chapter 8, p. Unit 8.12, Jun. 2007.

[53] A. Newman, “Product Review: The Precise World of Isotope Ratio Mass Spectrometry,” *Analytical Chemistry*, vol. 68, no. 11, p. 373A–377A, Jun. 1996.

[54] T. Kind and O. Fiehn, “Advances in structure elucidation of small molecules using mass spectrometry.,” *Bioanalytical reviews*, vol. 2, no. 1–4, pp. 23–60, Dec. 2010.

[55] C. Lifshitz and J. Laskin, *Principles of mass spectrometry applied to biomolecules*. John Wiley & Sons, Inc., 2006.

[56] M. Sharon, “How far can we go with structural mass spectrometry of protein complexes?,” *Journal of the American Society for Mass Spectrometry*, vol. 21, no. 4, pp. 487–500, Apr. 2010.

[57] C. Uetrecht, C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck, “High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 27, pp. 9216–20, Jul. 2008.

[58] A. El-Aneed, A. Cohen, and J. Banoub, “Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers,” *Applied Spectroscopy Reviews*, vol. 44, no. 3, pp. 210–230, Apr. 2009.

[59] J. J. Pitt, “Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry.,” *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*, vol. 30, no. 1, pp. 19–34, Feb. 2009.

[60] K. Downard, *Mass Spectrometry: A Foundation Course*. Royal Society of Chemistry, 2004, p. 210.

[61] J. H. Gross, *Mass Spectrometry: A Textbook*. Springer, 2011, p. 774.

[62] E. de Hoffmann and V. Stroobant, *Mass Spectrometry: Principles and Applications*. Wiley-Interscience, 2007, p. 502.

[63] D. G. Schmid, P. Grosche, H. Bandel, and G. Jung, “FTICR-mass spectrometry for high-resolution analysis in combinatorial chemistry.,” *Biotechnology and bioengineering*, vol. 71, no. 2, pp. 149–61.

[64] A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, and H. H. Hill, “Ion mobility-mass spectrometry.,” *Journal of mass spectrometry : JMS*, vol. 43, no. 1, pp. 1–22, Jan. 2008.

[65] N. F. Zinnel, P.-J. Pai, and D. H. Russell, “Ion mobility-mass spectrometry (IM-MS) for top-down proteomics: increased dynamic range affords increased sequence coverage.,” *Analytical chemistry*, vol. 84, no. 7, pp. 3390–7, Apr. 2012.

[66] C. Fenselau, “Tandem mass spectrometry: the competitive edge for pharmacology.,” *Annual review of pharmacology and toxicology*, vol. 32, pp. 555–78, Jan. 1992.

[67] A. W. Jones and H. J. Cooper, “Dissociation techniques in mass spectrometry-based proteomics.,” *The Analyst*, vol. 136, no. 17, pp. 3419–29, Sep. 2011.

[68] L. Sleno and D. A. Volmer, “Ion activation methods for tandem mass spectrometry.,” *Journal of mass spectrometry : JMS*, vol. 39, no. 10, pp. 1091–112, Oct. 2004.

[69] E. E. Jellen, A. M. Chappell, and V. Ryzhov, “Effects of size of noncovalent complexes on their stability during collision-induced dissociation.,” *Rapid communications in mass spectrometry : RCM*, vol. 16, no. 19, pp. 1799–804, Jan. 2002.

[70] J. M. Wells and S. a McLuckey, “Collision-induced dissociation (CID) of peptides and proteins.,” *Methods in enzymology*, vol. 402, no. 1993, pp. 148–85, Jan. 2005.

[71] P. Mayer, “The mechanisms of collisional activation of ions in mass spectrometry,” *Mass spectrometry reviews*, 2009.

[72] K. Vékey, “Internal Energy Effects in Mass Spectrometry,” *Journal of Mass Spectrometry*, vol. 31, no. 5, pp. 445–463, May 1996.

[73] S. A. McLuckey and D. E. Goeringer, “SPECIAL FEATURE:TUTORIAL Slow Heating Methods in Tandem Mass Spectrometry,” *Journal of Mass Spectrometry*, vol. 32, no. 5, pp. 461–474, May 1997.

[74] S. Yin and J. A. Loo, “Mass spectrometry detection and characterization of noncovalent protein complexes.,” *Methods in molecular biology (Clifton, N.J.)*, vol. 492, pp. 273–82, Jan. 2009.

[75] J. Sztáray, A. Memboeuf, L. Drahos, and K. Vékey, “Leucine enkephalin--a mass spectrometry standard.,” *Mass spectrometry reviews*, vol. 30, no. 2, pp. 298–320, 2011.

[76] F. Muntean and P. B. Armentrout, “Guided ion beam study of collision-induced dissociation dynamics: integral and differential cross sections,” *The Journal of Chemical Physics*, vol. 115, no. 3, p. 1213, Jul. 2001.

[77] O. Meroueh and W. L. Hase, “Energy transfer pathways in the collisional activation of peptides,” *International Journal of Mass Spectrometry*, vol. 201, no. 1–3, pp. 233–244, Jul. 2000.

[78] N. Rathore, T. A. Knotts IV, and J. J. de Pablo, “Density of states simulations of proteins,” *The Journal of chemical physics*, vol. 118, p. 4285, 2003.

[79] D. M. Black, A. H. Payne, and G. L. Glish, “Determination of cooling rates in a quadrupole ion trap.,” *Journal of the American Society for Mass Spectrometry*, vol. 17, no. 7, pp. 932–8, Jul. 2006.

[80] J. C. Smith, K. W. M. Siu, and S. P. Rafferty, “Collisional cooling enhances the ability to observe non-covalent interactions within the inducible nitric oxide synthase oxygenase domain: dimerization, complexation, and dissociation.,” *Journal of the American Society for Mass Spectrometry*, vol. 15, no. 5, pp. 629–38, May 2004.

[81] L. Drahos and K. Vékey, “MassKinetics: a theoretical model of mass spectra incorporating physical processes, reaction kinetics and mathematical descriptions,” *Journal of mass spectrometry*, no. 36, pp. 237–263, 2001.

[82] “Modelling energy transfer in molecular collisions: statistical theory versus experiment for highly excited toluene and azulene.”

[83] A. J. R. Heck, “Native mass spectrometry: a bridge between interactomics and structural biology.,” *Nature methods*, vol. 5, no. 11, pp. 927–33, Nov. 2008.

[84] J. Lengqvist, A. Mata de Urquiza, T. Perlmann, J. Sjövall, and W. J. Griffiths, “Specificity of receptor-ligand interactions and their effect on dimerisation as observed by electrospray mass spectrometry: bile acids form stable adducts to the RXRalpha.,” *Journal of mass spectrometry : JMS*, vol. 40, no. 11, pp. 1448–61, Nov. 2005.

[85] T. L. Pukala, B. T. Ruotolo, M. Zhou, A. Politis, R. Stefanescu, J. A. Leary, and C. V Robinson, “Subunit architecture of multiprotein assemblies determined using restraints from gas-phase measurements.,” *Structure (London, England : 1993)*, vol. 17, no. 9, pp. 1235–43, Sep. 2009.

[86] H. J. Maple, R. A. Garlish, L. Rigau-Roca, J. Porter, I. Whitcombe, C. E. Prosser, J. Kennedy, A. J. Henry, R. J. Taylor, M. P. Crump, and J. Crosby, “Automated protein-ligand interaction screening by mass spectrometry.,” *Journal of medicinal chemistry*, vol. 55, no. 2, pp. 837–51, Jan. 2012.

[87] C. Bich, M. Scott, A. Panagiotidis, R. J. Wenzel, A. Nazabal, and R. Zenobi, “Characterization of antibody-antigen interactions: comparison between surface plasmon resonance measurements and high-mass matrix-assisted laser desorption/ionization mass

spectrometry.," *Analytical biochemistry*, vol. 375, no. 1, pp. 35–45, Apr. 2008.

[88] M. Karas, U. Bahr, and T. Dülcks, "Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine," *Fresenius' Journal of Analytical Chemistry*, vol. 366, no. 6–7, pp. 669–676, Mar. 2000.

[89] Z. Takáts, J. M. Wiseman, B. Gologan, and R. G. Cooks, "Electrosonic spray ionization. A gentle technique for generating folded proteins and protein complexes in the gas phase and for studying ion-molecule reactions at atmospheric pressure.," *Analytical chemistry*, vol. 76, no. 14, pp. 4050–8, Jul. 2004.

[90] C. V Robinson, A. Sali, and W. Baumeister, "The molecular sociology of the cell.," *Nature*, vol. 450, no. 7172, pp. 973–82, Dec. 2007.

[91] K. Lorenzen, A. Vannini, P. Cramer, and A. J. R. Heck, "Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture.," *Structure (London, England : 1993)*, vol. 15, no. 10, pp. 1237–45, Oct. 2007.

[92] J. M. Daniel, S. D. Friess, S. Rajagopalan, S. Wendt, and R. Zenobi, "Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry," *International Journal of Mass Spectrometry*, vol. 216, no. 1, pp. 1–27, Apr. 2002.

[93] P. M. Mayer and E. Martineau, "Gas-phase binding energies for non-covalent Aβ-40 peptide/small molecule complexes from CID mass spectrometry and RRKM theory.," *Physical chemistry chemical physics : PCCP*, vol. 13, no. 11, pp. 5178–86, Mar. 2011.

[94] V. J. Nesatyy, "Gas-phase binding of non-covalent protein complexes between bovine pancreatic trypsin inhibitor and its target enzymes studied by electrospray ionization tandem mass spectrometry.," *Journal of mass spectrometry : JMS*, vol. 36, no. 8, pp. 950–9, Aug. 2001.

[95] M. Schäfer, C. Schmuck, M. Heil, H. J. Cooper, C. L. Hendrickson, M. J. Chalmers, and A. G. Marshall, "Determination of the activation energy for unimolecular dissociation of a non-covalent gas-phase peptide: substrate complex by infrared multiphoton dissociation fourier transform ion cyclotron resonance mass spectrometry.," *Journal of the American Society for Mass Spectrometry*, vol. 14, no. 11, pp. 1282–9, Nov. 2003.

[96] J. L. P. Benesch and C. V Robinson, "Mass spectrometry of macromolecular assemblies: preservation and dissociation.," *Current opinion in structural biology*, vol. 16, no. 2, pp. 245–51, Apr. 2006.

[97] R. H. H. van den Heuvel and A. J. R. Heck, "Native protein mass spectrometry: from intact oligomers to functional machineries.," *Current opinion in chemical biology*, vol. 8, no. 5, pp. 519–26, Oct. 2004.

[98] T. Meyer, X. de la Cruz, and M. Orozco, "An atomistic view to the gas phase proteome.," *Structure (London, England : 1993)*, vol. 17, no. 1, pp. 88–95, Jan. 2009.

[99] P. Kebarle and M. Peschke, "On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions," *Analytica Chimica Acta*, vol. 406, no. 1, pp. 11–35, Feb. 2000.

[100] M. Z. Steinberg, K. Breuker, R. Elber, and R. B. Gerber, "The dynamics of water evaporation from partially solvated cytochrome c in the gas phase.," *Physical chemistry chemical physics : PCCP*, vol. 9, no. 33, pp. 4690–7, Sep. 2007.

[101] M. Z. Steinberg, R. Elber, F. W. McLafferty, R. B. Gerber, and K. Breuker, "Early structural evolution of native cytochrome c after solvent removal.," *Chembiochem : a European journal of chemical biology*, vol. 9, no. 15, pp. 2417–23, Oct. 2008.

[102] G. Verbeck, W. Hoffmann, and B. Walton, "Soft-landing preparative mass spectrometry.," *The Analyst*, vol. 137, no. 19, pp. 4393–407, Oct. 2012.

[103] H. Oh, K. Breuker, S. K. Sze, Y. Ge, B. K. Carpenter, and F. W. McLafferty, "Secondary and

tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 25, pp. 15863–8, Dec. 2002.

[104] K. Barylyuk, R. M. Balabin, D. Grünstein, R. Kikkeri, V. Frankevich, P. H. Seeberger, and R. Zenobi, "What happens to hydrophobic interactions during transfer from the solution to the gas phase? The case of electrospray-based soft ionization methods.," *Journal of the American Society for Mass Spectrometry*, vol. 22, no. 7, pp. 1167–77, Jul. 2011.

[105] R. Frański, B. Gierczyk, G. Schroeder, M. Frańska, and B. Wyrwas, "Do hydrophobic interactions exist in the gas phase?," *Rapid communications in mass spectrometry : RCM*, vol. 22, no. 8, pp. 1339–43, Apr. 2008.

[106] K. Breuker and F. W. McLafferty, "The thermal unfolding of native cytochrome c in the transition from solution to gas phase probed by native electron capture dissociation.," *Angewandte Chemie (International ed. in English)*, vol. 44, no. 31, pp. 4911–4, Aug. 2005.

[107] K. Breuker, H. Oh, D. M. Horn, B. A. Cerda, and F. W. McLafferty, "Detailed Unfolding and Folding of Gaseous Ubiquitin Ions Characterized by Electron Capture Dissociation," *Journal of the American Chemical Society*, vol. 124, no. 22, pp. 6407–6420, Jun. 2002.

[108] J. A. Loo, "Studying noncovalent protein complexes by electrospray ionization mass spectrometry.," *Mass spectrometry reviews*, vol. 16, no. 1, pp. 1–23.

[109] W. Wang, E. N. Kitova, and J. S. Klassen, "Bioactive recognition sites may not be energetically preferred in protein-carbohydrate complexes in the gas phase.," *Journal of the American Chemical Society*, vol. 125, no. 45, pp. 13630–1, Nov. 2003.

[110] K. Breuker, "The study of protein–ligand interactions by mass spectrometry—a personal view," *International Journal of Mass Spectrometry*, vol. 239, no. 1, pp. 33–41, Dec. 2004.

[111] A. Mittermaier and L. E. Kay, "New tools provide new insights in NMR studies of protein dynamics.," *Science (New York, N.Y.)*, vol. 312, no. 5771, pp. 224–8, Apr. 2006.

[112] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, and D. Kern, "Intrinsic motions along an enzymatic reaction trajectory.," *Nature*, vol. 450, no. 7171, pp. 838–44, Dec. 2007.

[113] O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot, "Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution.," *Science (New York, N.Y.)*, vol. 320, no. 5882, pp. 1471–5, Jun. 2008.

[114] P. Tuffery and P. Derreumaux, "Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches.," *Journal of the Royal Society, Interface / the Royal Society*, vol. 9, no. 66, pp. 20–33, Jan. 2012.

[115] C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky, and A. K. Dunker, "Comparing and combining predictors of mostly disordered proteins.," *Biochemistry*, vol. 44, no. 6, pp. 1989–2000, Feb. 2005.

[116] L.-W. Yang, a J. Rader, X. Liu, C. J. Jursa, S. C. Chen, H. a Karimi, and I. Bahar, "oGNM: online computation of structural dynamics using the Gaussian Network Model.," *Nucleic acids research*, vol. 34, no. Web Server issue, pp. W24–31, Jul. 2006.

[117] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar, "PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique," *Journal of Chemical Theory and Computation*, vol. 1, no. 6, pp. 1304–1311, Nov. 2005.

[118] D. Motiejunas, R. Gabdoulline, T. Wang, A. Feldman-Salit, T. Johann, P. J. Winn, and R. C. Wade, "Protein-protein docking by simulating the process of association subject to

biochemical constraints.," *Proteins*, vol. 71, no. 4, pp. 1955–69, Jun. 2008.

[119] A. G. Palmer, "NMR characterization of the dynamics of biomacromolecules.," *Chemical reviews*, vol. 104, no. 8, pp. 3623–40, Aug. 2004.

[120] J. W. Peng, V. Thanabal, and G. Wagner, "2D heteronuclear NMR measurements of spin-lattice relaxation times in the rotating frame of X nuclei in heteronuclear HX spin systems," *Journal of Magnetic Resonance (1969)*, vol. 94, no. 1, pp. 82–100, Aug. 1991.

[121] J. W. Peng and G. Wagner, "Mapping of the spectral densities of nitrogen-hydrogen bond motions in Eglin c using heteronuclear relaxation experiments," *Biochemistry*, vol. 31, no. 36, pp. 8571–8586, Sep. 1992.

[122] R. Ishima and K. Nagayama, "Protein Backbone Dynamics Revealed by Quasi Spectral Density Function Analysis of Amide N-15 Nuclei," *Biochemistry*, vol. 34, no. 10, pp. 3162–3171, Mar. 1995.

[123] N. Farrow, O. Zhang, A. Szabo, D. Torchia, and L. Kay, "Spectral density function mapping using 15N relaxation data exclusively," *Journal of Biomolecular NMR*, vol. 6, no. 2, Sep. 1995.

[124] D. Fushman, N. Tjandra, and D. Cowburn, "to direct determination of protein dynamics from 15N NMR relaxation at multiple fields, independent of variable 15N chemical shift anisotropy and chemical exchange," *Journal of the American Chemical Society*, vol. 121, pp. 8577–8582, 1999.

[125] D. Fushman, N. Tjandra, and D. Cowburn, "Direct Measurement of 15 N Chemical Shift Anisotropy in Solution," *Journal of the American Chemical Society*, vol. 120, no. 42, pp. 10947–10952, Oct. 1998.

[126] I. Q. Phan, J. Boyd, and I. D. Campbell, "Dynamic studies of a fibronectin type I module pair at three frequencies: Anisotropic modelling and direct determination of conformational exchange.," *Journal of biomolecular NMR*, vol. 8, no. 4, pp. 369–78, Dec. 1996.

[127] C. L. Perrin and T. J. Dwyer, "Application of two-dimensional NMR to kinetics of chemical exchange," *Chemical Reviews*, vol. 90, no. 6, pp. 935–967, Sep. 1990.

[128] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins.," *Science (New York, N.Y.)*, vol. 330, no. 6002, pp. 341–6, Oct. 2010.

[129] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity," *Journal of the American Chemical Society*, vol. 104, no. 17, pp. 4546–4559, Aug. 1982.

[130] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results," *Journal of the American Chemical Society*, vol. 2, no. 1, 1982.

[131] G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn, "Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins," *Journal of the American Chemical Society*, vol. 112, no. 12, pp. 4989–4991, Jun. 1990.

[132] D. I. Freedberg, Y.-X. Wang, S. J. Stahl, J. D. Kaufman, P. T. Wingfield, Y. Kiso, and D. A. Torchia, "Flexibility and Function in HIV Protease: Dynamics of the HIV-1 Protease Bound to the Asymmetric Inhibitor Kynostatin 272 (KNI-272)," *Journal of the American Chemical Society*, vol. 120, no. 31, pp. 7916–7923, Aug. 1998.

[133] P. Luginbühl and K. Wüthrich, "Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules," *Progress in Nuclear Magnetic Resonance Spectroscopy*, no. 40, pp. 199–247, 2002.

[134] N. Tjandra, S. E. Feller, R. W. Pastor, and A. Bax, “Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation,” *Journal of the American Chemical Society*, vol. 117, no. 50, pp. 12562–12566, Dec. 1995.

[135] D. S. Korchuganov, I. E. Gagnidze, E. N. Tkach, A. a Schulga, M. P. Kirpichnikov, and A. S. Arseniev, “Determination of protein rotational correlation time from NMR relaxation data at various solvent viscosities.,” *Journal of biomolecular NMR*, vol. 30, no. 4, pp. 431–42, Dec. 2004.

[136] P. Dosset, J. C. Hus, M. Blackledge, and D. Marion, “Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data.,” *Journal of biomolecular NMR*, vol. 16, no. 1, pp. 23–8, Jan. 2000.

[137] D. Fushman, S. Cahill, and D. Cowburn, “The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration.,” *Journal of molecular biology*, vol. 266, no. 1, pp. 173–94, Feb. 1997.

[138] J. Hall and D. Fushman, “Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G,” *Journal of biomolecular NMR*, vol. 27, no. 3, pp. 261–275, 2003.

[139] J. B. Hall and D. Fushman, “Variability of the 15N chemical shielding tensors in the B3 domain of protein G from 15N relaxation measurements at several fields. Implications for backbone order parameters.,” *Journal of the American Chemical Society*, vol. 128, no. 24, pp. 7855–70, Jun. 2006.

[140] A. Hoeben, B. Landuyt, M. S. Highley, H. Wildiers, A. T. Van Oosterom, and E. A. De Bruijn, “Vascular endothelial growth factor and angiogenesis.,” *Pharmacological reviews*, vol. 56, no. 4, pp. 549–80, Dec. 2004.

[141] S. Iyer and K. R. Acharya, “Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines.,” *The FEBS journal*, vol. 278, no. 22, pp. 4304–22, Nov. 2011.

[142] W. J. Fairbrother, M. A. Champe, H. W. Christinger, B. A. Keyt, and M. A. Starovasnik, “1H, 13C, and 15N backbone assignment and secondary structure of the receptor-binding domain of vascular endothelial growth factor.,” *Protein science : a publication of the Protein Society*, vol. 6, no. 10, pp. 2250–60, Oct. 1997.

[143] Y. a Muller, H. W. Christinger, B. a Keyt, and a M. de Vos, “The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 A resolution: multiple copy flexibility and receptor binding.,” *Structure (London, England : 1993)*, vol. 5, no. 10, pp. 1325–38, Oct. 1997.

[144] Y. a Muller, B. Li, H. W. Christinger, J. a Wells, B. C. Cunningham, and a M. de Vos, “Vascular endothelial growth factor: crystal structure and functional mapping of the kinase domain receptor binding site.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 14, pp. 7192–7, Jul. 1997.

[145] B. A. C. Horta, J. J. V. Cirino, and R. B. de Alencastro, “On the structure, interactions, and dynamics of bound VEGF.,” *Journal of molecular graphics & modelling*, vol. 26, no. 7, pp. 1091–103, Apr. 2008.

[146] N. Ferrara, “The Biology of Vascular Endothelial Growth Factor,” *Endocrine Reviews*, vol. 18, no. 1, pp. 4–25, Feb. 1997.

[147] L. Coultas, K. Chawengsaksophak, and J. Rossant, “Endothelial cells and VEGF in vascular development.,” *Nature*, vol. 438, no. 7070, pp. 937–45, Dec. 2005.

[148] G. Neufeld, T. Cohen, S. Gengrinovitch, and Z. Poltorak, “Vascular endothelial growth factor (VEGF) and its receptors,” *FASEB J*, vol. 13, no. 1, pp. 9–22, Jan. 1999.

[149] N. Lu, Y. Gao, Y. Ling, Y. Chen, Y. Yang, H.-Y. Gu, Q. Qi, W. Liu, X.-T. Wang, Q.-D. You, and Q.-L. Guo, “Wogonin suppresses tumor growth in vivo and VEGF-induced angiogenesis through inhibiting tyrosine phosphorylation of VEGFR2.,” *Life sciences*, vol. 82, no. 17–18, pp. 956–63, Apr. 2008.

[150] W. D. Brazelle, W. Shi, and D. W. Siemann, “VEGF-associated tyrosine kinase inhibition increases the tumor response to single and fractionated dose radiotherapy.,” *International journal of radiation oncology, biology, physics*, vol. 65, no. 3, pp. 836–41, Jul. 2006.

[151] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, “Vascular endothelial growth factor (VEGF) inhibition--a critical review.,” *Anti-cancer agents in medicinal chemistry*, vol. 7, no. 2, pp. 223–45, Mar. 2007.

[152] K. A. Krishna, G. V. Rao, and K. R. S. S. Rao, “Chaperonin GroEL: structure and reaction cycle.,” *Current protein & peptide science*, vol. 8, no. 5, pp. 418–25, Oct. 2007.

[153] P. B. Sigler, Z. Xu, H. S. Rye, S. G. Burston, W. A. Fenton, and A. L. Horwich, “Structure and function in GroEL-mediated protein folding.,” *Annual review of biochemistry*, vol. 67, pp. 581–608, Jan. 1998.

[154] H. Taguchi, “Chaperonin GroEL meets the substrate protein as a ‘load’ of the rings.,” *Journal of biochemistry*, vol. 137, no. 5, pp. 543–9, May 2005.

[155] J. MONOD, J. WYMAN, and J. P. CHANGEUX, “ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL.,” *Journal of molecular biology*, vol. 12, pp. 88–118, May 1965.

[156] D. E. Koshland, G. Némethy, and D. Filmer, “Comparison of experimental binding data and theoretical models in proteins containing subunits.,” *Biochemistry*, vol. 5, no. 1, pp. 365–85, Jan. 1966.

[157] D. Rivenzon-Segal, S. G. Wolf, L. Shimon, K. R. Willison, and A. Horovitz, “Sequential ATP-induced allosteric transitions of the cytoplasmic chaperonin containing TCP-1 revealed by EM analysis.,” *Nature structural & molecular biology*, vol. 12, no. 3, pp. 233–7, Mar. 2005.

2.

Molecular recognition at protein surface in solution and gas phase: five VEGF peptidic ligands show inverse affinity when studied by NMR and CID-MS

Molecular Recognition at Protein Surface in Solution and Gas Phase: Five VEGF Peptidic Ligands Show Inverse Affinity When Studied by NMR and CID-MS

Andrey Dyachenko,[1] Michael Goldflam,[1] Marta Vilaseca,[1] Ernest Giralt[1,2]
[1] *Institute for Research in Biomedicine, Parc Científic de Barcelona, Barcelona, Spain*

[2] *Department of Organic Chemistry, Universidad de Barcelona, Barcelona, Spain*

Received 15 February 2010; revised 18 March 2010; accepted 18 March 2010
Published online 4 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.21462

ABSTRACT:

Protein-protein interactions comprise of collection of molecular recognition events that take place at protein surfaces. A better understanding of the mechanism behind these interactions would provide deeper insight into the nature of many diseases, caused by the malfunction of protein networks, and contribute to design of molecules for efficient modulating of these interactions. One major factor in molecular recognition mechanism is interaction of reacting species with aqueous media. Thus, comparative study of noncovalent complex behavior in solution and gas phase can provide valuable information about the role of the solvent. Here examined interactions of vascular endothelial growth factor (VEGF) protein with five peptidic ligands of the same molecular weight but with different affinities. Interactions of VEGF with ligands in solution were studied by ITC and NMR, and K_D*s were determined. Gas phase stability was addressed using CID-MS approach. The energy transfer model was taken and adapted for the calculation of binding energy. Peptides were ranked on the basis of both solution and gas phase affinity to VEGF. The results indicate that the ranking of peptides in terms of affinity in solution is reversed compared with the gas phase ranking. This observation opens up a vast field for the future study of the system, and the determination and characterization of factors, responsible for the change of stability of noncovalent protein-ligand complexes upon complete or partial removal of the solvent. © 2010 Wiley Periodicals, Inc. Biopolymers (Pept Sci) 94: 689–700, 2010.*

Keywords: *protein-ligand interactions; molecular recognition; noncovalent interactions; gas-phase interactions; mass spectrometry; vascular endothelial growth factor (VEGF)*

INTRODUCTION

One of the major highlights from biological sciences in this last decade has undoubtedly been the discovery that proteins inside the cells have a very rich "social life."[1] They rarely work alone but interact with several others to form multi-protein complexes. These complexes, which act as molecular machines, are in charge of most of the multiple functions of proteins, such as catalysis, transport, and signal transduction. In general, protein-protein networks are topologically heterogeneous in the sense that a few highly connected proteins-protein

Additional Supporting Information may be found in the online version of this article.
Correspondence to: Ernest Giralt; e-mail: ernest.giralt@irbbarcelona.org
Contract grant sponsor: MCYT-FEDER
Contract grant number: Bio2008-00799
Contract grant sponsors: *Generalitat de Catalunya* (XRB and *Grup Consolidat*)

hubs mediate interactions between numerous less connected proteins. The number of proteins per node varies greatly, ranging from two or three to more than twenty. From a dynamic point of view, the interactions vary from very transient to highly stable.[2–6]

The formation of holes or knots in the protein network of a healthy cell normally leads to disease.[1] By a hole, we mean the loss-of-function of a protein hub. p53 is a typical example of such a hub. Loss of function of p53 is strongly associated with cancer. In fact, mutations of the gene coding for p53 have been identified in more than 60% of cancer patients. We talk about a knot in the net when a series of new and unwanted protein-protein interactions appears. Protein self-assembly is a special case—"narcissistic"—of protein-protein interactions.[7–9] In this regard amyloid diseases, such as Alzheimer's or Parkinson's, can also be considered to derive from the formation of knots in the net. In recent years we have been working on the design of therapeutic strategies that modulate this socially "impolite" behavior. Thus, we have developed cationic calixarenes able to rescue the structure and stability of mutated versions of p53.[10,11] We have also designed peptides and peptide-decorated gold nanoparticles that interfere with the formation of the β-amyloid aggregates associated with Alzheimer's disease.[12–16]

Protein-protein interactions are the result of an ensemble of exquisitely regulated molecular recognition events that take place at protein surfaces. This can be referred to as a "protein recognition code." To understand protein-protein interactions and to achieve the efficient design of molecules with the capacity to modulate these protein-protein interactions, it is necessary to decipher this molecular recognition code, the language that proteins use to communicate. Unfortunately, progress in this field is highly unsatisfactory. Indeed, we are not completely illiterate, in the sense that we know the letters of this alphabet. They are the noncovalent interactions, such as hydrogen bonds, electrostatic interactions, π-cation interactions, Van der Waals forces, and the others. However, we could be compared with a child who is learning to read and attempts Dickens's Oliver Twist.

Where do our difficulties arise from? As pointed out by the seminal work of D. H. Williams,[17,18] G. Whitesides,[19] and M. Gilson,[20] dynamics—flexibility—and solvent effects are the two villains in this field. Equation 1 represents two molecules, A and B, that form a noncovalent complex AB. Either A and B are proteins, or one is a protein and the other a ligand (for example a peptide).

$$A + B \Leftrightarrow A ::: B \qquad (1)$$

Contract grant sponsor:

The difference with a conventional chemical reaction is that no covalent bond formation is involved. In contrast, A and B are held together by many noncovalent bonds. The Gibbs energy in aqueous solution (ΔG) of this process can be decomposed in several terms[18]:

$$\Delta G = \Delta G_{\mathrm{t+r}} + n\Delta G_{\mathrm{r}} + \mathrm{A}\Delta G_{\mathrm{h}} + \Sigma\Delta G_{\mathrm{p}}$$

where $\Delta G_{\mathrm{t+r}}$ (in fact, $\Delta S_{\mathrm{t+r}}$) corresponds to the loss of translational and rotational motion as a result of the transformation of two molecular species (A and B) in one molecular species (A:::B); $n\Delta G_{\mathrm{r}}$ (in fact, $n\Delta S_{\mathrm{r}}$) is the entropic cost associated with restricted rotations upon binding; ΔG_{r} is the entropic cost for each restricted rotation and n is the number of restricted rotations. $A\Delta G_{\mathrm{h}}$ (in fact, $A\Delta S_{\mathrm{h}}$ again) is the term dealing with the hydrophobic effect, where A is the area buried in the binding site and inaccessible to water and ΔG_{h} is the Gibbs energy of hydrophobic effect per unit area. Finally, $\Sigma\Delta G_{\mathrm{p}}$ is the sum of the Gibbs energies of all the polar interactions in the complex, where each ΔG_{p} term has an enthalpic (ΔH_{p}) and an entropic (ΔS_{p}) component.

Of all these terms, only $n\Delta G_{\mathrm{r}}$ and $A\Delta G_{\mathrm{h}}$ can be estimated in a relatively accurate way. In contrast, estimation—even a rough one—of $\Delta G_{\mathrm{t+r}}$ or $\Sigma\Delta G_{\mathrm{p}}$ presents severe problems. Estimation of $\Delta G_{\mathrm{t+r}}$ can at first glance appear to be straightforward, namely, a simple calculation of the entropic loss associated with passing from two free rigid bodies to a single rigid body. Nevertheless, things are not quite so simple, and true $\Delta G_{\mathrm{t+r}}$ values are probably at least one order of magnitude smaller. This may be because the non-covalent interactions holding A and B together are in fact very weak. Thus, at room temperature one can expect kinetic behavior with continuous formation and breakage of noncovalent bonds. This transient bonding scheme could be responsible for this diminution of $\Delta G_{\mathrm{t+r}}$. In other words, the A:::B complex cannot be treated as a rigid body. Accurate estimations of enthalpic and entropic components of $\Sigma\Delta G_{\mathrm{p}}$ are hindered by difficulties in the following: to estimate the gain in entropy caused by the release of solvating water upon binding; to estimate changes of both entropy and enthalpy as a result of the release of counter-ions; and to estimate the effective dielectric permittivity in water-accessible regions of proteins. As discussed below, modern mass spectrometry methods have opened the door to the quantitative study of binding in gas phase. Comparison between binding in water and in gas phase appears as a promising approach shad light on this difficult field.

Peptide molecules are wonderful tools with which to examine protein-protein interactions[21,22] and to modulate them.[12,13] Here we present a comparative study of the interaction of a family of diastereomeric cyclic peptides and a therapeutically relevant protein, the Vascular Endothelial Growth Factor A (VEGF-A).

VEGF-A is a member of the platelet-derived growth factor (PDGF)/vascular endothelial growth factor (VEGF) family, which comprises seven members (VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGF-E, VEGF-F, and PlGF). VEGF-A is a globular water-soluble covalent homodimer, in which monomers are connected by two disulfide bridges.[23] This growth factor is expressed in the human body in at least seven isoforms, containing 121, 145, 148, 165, 183, 189, and 206 residues per monomer.[24]

The role of VEGF-A as a therapeutic target is extensively described in a number of excellent reviews (see, for example, Hoeben et al.[24] or Ferrara and Kerbel[25]). VEGF-A is a major regulator of normal and abnormal angiogenesis, including that associated with tumors and several intraocular syndromes.[26] It binds to two receptor tyrosine kinases (RTK), VEGFR-1 (Flt-1), and VEGFR-2 (KDR, Flk-1). In particular, VEGFR-2 is the major mediator of the mitogenic, angiogenic and permeability-enhancing effects of VEGF-A.

Generally, VEGF-A up-regulation is associated with tumor growth, facilitating the spread of blood vessels and blood supply of the tumor. Inhibition of VEGF-induced angiogenesis hinders this growth[27] and increase the tumor response to radiotherapy.[28] In addition to tumor growth VEGF-A is involved in number of pathological conditions, including cardiovascular diseases, diabetic retinopathy, rheumatoid arthritis and psoriasis. Furthermore, VEGF-A is required for the normal functioning and development of mammals, thus making it an attractive target for both antiangiogenic and proangiogenic therapy.

Here we used a 11-109 construct of VEGF-A$_{121}$ (later in the text referred to as VEGF). Although truncated, it was shown to retain its biological activity, because it still contains the VEGFR-1 and VEGFR-2 binding domain.[29] Thus, it can be used for the identification of inhibitors of VEGF-A:::VEGFR-1 and VEGF-A:::VEGFR-2 interactions.

Phage-display-derived cyclic peptide,[23] binds to VEGF in 1:2 stoichiometry with a dissociation constant (K_D) of approx. 1 μM, overlapping to a large extent with the VEGFR-1 binding site. It is a 2340 Da 19 amino acid peptide (**P-wt** in Table I), stabilized with a C5-C15 disulfide bond. It is poorly ordered in the unbound state, but in complex with VEGF it adopts a strongly amphipathic secondary conformation, positioning hydrophobic residues towards VEGF binding site and exposing hydrophilic residues to the solvent. The NMR structure of VEGF in complex with **P-wt** (1KAT in PDB database) is shown in Figure 1.

A complex of VEGF with **P-wt** was chosen as a model for this study. To gain further insight into the nature of noncovalent interactions, we synthesized 5 peptide ligands to VEGF (Table I). Each ligand is a **P-wt** peptide with one L amino acid substituted with the D analog. The amino acids were selected on the basis of an Ala scan of the binding area of the ligand, performed by Pan et al.[23] (See Table I). Although substitution of the same amino acid with Ala and with D enantiomer should have a different effect on the structure of peptide, Ala scan can provide a rough estimation of the relative importance of given amino acid side chain for binding.[30] The following four amino acids were chosen for the substitution: Arg18, Ile7, Met10, and Phe16 (Table I).

Table I Synthesized Cyclic Peptides, Arranged by the Predicted Decrease of Affinity to VEGF (From Top to Bottom)

Ligand	Sequence[a,b,c]	P-wt/P-mut[d]
P-wt	GGNECDIARMWEWECFERL	1
P-18r	GGNECDIARMWEWECFE**r**L	NA
P-7i	GGNECD**i**ARMWEWECFERL	25
P-10m	GGNECDIAR**m**WEWECFERL	200
P-16f	GGNECDIARMWEWEC**f**ERL	>2000

[a] The lowercase bold letter represents the D amino acid.

[b] All the peptides were cyclized with C5-C15 disulfide bridge.

[c] All the peptides have a carboxamide group at their C-termini.

[d] Decrease of binding strength after substitution of the 18th, 7th, 10th, and 16th amino acid residue with Ala. Data is taken from Ref. 23.

The purpose of designing the D analogs of peptides was to introduce a local perturbation into the structure of the ligand, leaving as many of its properties as possible intact. The molecular weight of the ligand did not change, which simplifies the interpretation of mass spectrometry data, and no new chemical groups were introduced. No rotatable bonds were added or removed, thus leaving the number of degrees of freedom of the ligand unchanged. However, the perturbation was strong enough to cause notable changes in the strength of VEGF-ligand interactions.

Chemical Shift Perturbation Nuclear Magnetic Resonance (CSP NMR) is a powerful technique that allows the mapping of protein-protein and protein-ligand interactions[31] and the determination of the respective K_Ds.[32] To determine the K_D, protein is titrated with a ligand (or vice versa) and the HSQC spectrum is recorded for each point. Depending on the labeling and acquisition strategy, changes in the protein or ligand spectrum can be monitored. Interactions induce an alteration in the chemical shift of certain peaks ($\Delta\delta$), which can be correlated with the fraction of observed species in the bound state. Plotting $\Delta\delta$ versus the concentration of titrant and fitting the resulting plot to the respective model provides the K_D value.

Mass spectrometry (MS) is a well-established method that allows the study of species in gas phase, including small organic molecules,[33] large biomolecules,[34] and even intact viruses.[35] In particular, the development of so called "soft ionization" methods, Electrospray Ionization (ESI) and Matrix Assisted Laser Desorption-Ionization (MALDI), capable

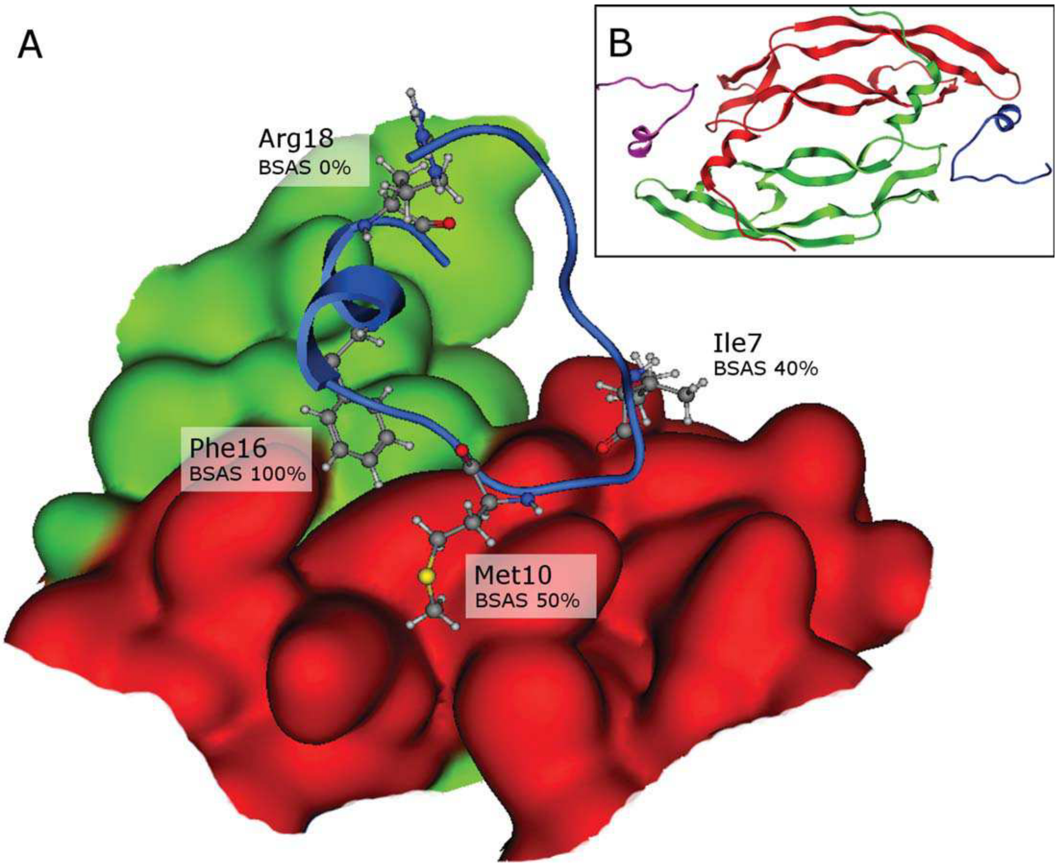


FIGURE 1 P-wt peptide in complex with 11-109 construct of VEGF$_{121}$. A: **P-wt** in the binding pocket. Amino acid residues, substituted by D-enantiomers in the present work, are shown as ball-and-stick models, with buried solvent accessible surface specified. B: 3D view of VEGF$_{121}$ dimer in complex with **P-wt** peptide (1KAT PDB entry).

of ionizing intact protein complexes, has facilitated the extensive study of noncovalent interactions by MS.[36–38]

Several features of MS make it an attractive approach to address biological molecules. First, it is a sensitive method, with the capacity to detect the samples at submicromolar concentrations, and it requires, in the best of cases, several nanograms of the sample for a single run. This spectacularly low sample consumption of the method makes it highly suited for the screening of large libraries of compounds.[39] Another feature of the method is mass accuracy, which can reach about 1 ppm for the large molecules, thereby allowing straightforward assignment of the signals. Also, the method is fast and, in many cases, it can be automated.[36] Finally, the critical feature in the context of the present work is that MS allows observations of biomolecules without solvation shells, because species are studied in the gas phase. Transfer from solution to the gas phase affects structure and interaction forces of molecules. However, this gives valuable information about the behavior of the system unaffected by the solvent.

Collision induced dissociation (CID) is a "slow-heating" dissociation method, used to provide the ion with enough energy to fragment (brake covalent bonds) or dissociate (break noncovalent bonds).[40] The ion is "heated" while moving through the collision cell, which is filled with inert gas, such as He, N_2, Ar, Xe or CO_2. Impacts with gas molecules provide an ion with excess internal energy, which is used for dissociation. For more details, see the extensive review by Mayer et al.[41]

RESULTS

Chemical Shift Perturbation NMR

The affinity of each of the five ligands to VEGF in solution was determined with NMR by analyzing chemical shift perturbations (CSP). Sample, containing VEGF in 100 μM concentration, was titrated with a ligand. Concentrations of ligands were chosen so that the optimal range of titration curve was covered, on the basis of the expected values of re-

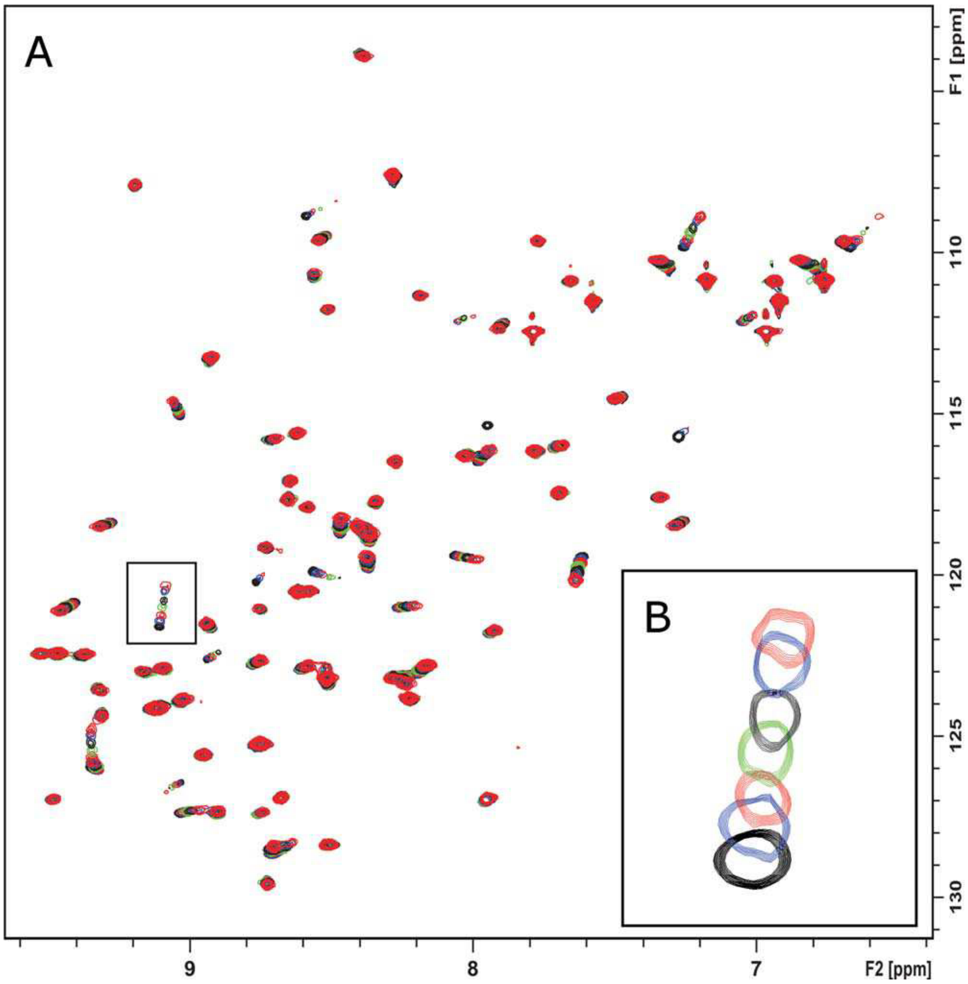


FIGURE 2 Determination of K_D of VEGF:::**P-7i** complex. A: ^{15}N-^{1}H HSQC spectra of a 100 μM sample of (methyl ^{13}C)-Met-all-^{15}N VEGF$_{11\text{-}109}$ at 45°C titrated with **P-7i** (600 MHz with cryoprobe). B: Zoom of Lys48 shifts.

spective K_Ds. ^{15}N-^{1}H HSQC spectrum was recorded for each titration point during 40 min (see Figure 2). For each ligand, four peaks with the strongest displacement of chemical shift ($\Delta\delta$) were used for the K_D calculations. The curves for $\Delta\delta$ versus total ligand concentration in the sample ($[L_0]$) were fitted to the analytical model, assuming that two binding sites are equal and independent [Eq. (2)].

$$\frac{[L_0]-[L]}{[P_0]} = \frac{K_D + [L_0] + 2[P_0] - \sqrt{(K_D + [L_0])^2 + 4[P_0](K_D - [L_0] + [P_0])}}{2} \tag{2}$$

where $[P_0]$ is the total protein concentration and $[L]$ is the concentration of unbound ligand in solution (see Figure 3). The fact that R^2 values for most of the fittings were better than 0.99 confirms that this model provides the acceptable description of the system. The average value of K_D was taken for each complex, and ligands were ranked on the basis of their affinity to VEGF in solution (Table II).

CID Mass Spectrometry

All the MS experiments were performed using a Waters SYNAPT HDMS Time-of-flight (TOF) mass spectrometer. The exact mass of the recombinant protein construct was determined under denaturating conditions (H_2O/CH_3CN/Acetic Acid 49/49/2). Convolution of the nanoESI spectra revealed the exact mass of VEGF to be 23520 ± 10 Da, which is in good agreement with the calculated mass, considering the fact that the protein was ^{15}N labeled for the NMR purposes.

Aqueous ammonium acetate (AmAc) solution is the buffer most widely used for the MS experiments with noncovalent complexes.[42] This is because of the volatility of its components and its near-physiological pH. These features allow the native structure of many proteins to remain intact, and, in many

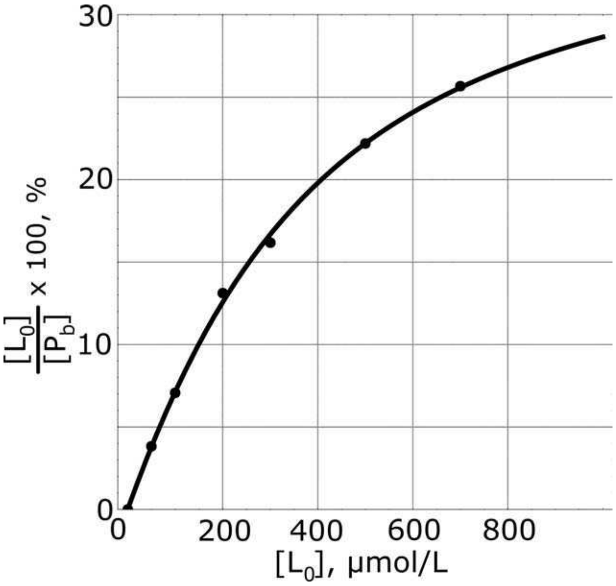


FIGURE 3 Relative amount of bound ligand plotted against the total ligand concentration and fitted to the "2 independent equal binding sites" model. Lys48 peak displacement in VEGF:::**P-7i** complex spectrum was used.

cases, to preserve noncovalent interactions. Concentrations of AmAc up to 1*M* for noncovalent complexes[42] and up to 8*M* for single proteins in the native state[43] have been used without negative impact on the spectrum. Moreover, in case of the presence of nonvolatile adducts in the sample solution, addition of AmAc can substantially reduce the broadening of the peaks, presumably, by replacing of adduct ions, such as Na^+, with ammonium ions NH_4^+.[43,44]

In this study, we used BioRad BioSpin-6 columns two times for each sample for effective desalting/buffer exchange, which resulted in very low concentration of Na^+ ions in the sample. Thus, high concentrations of AmAc were not required. Nevertheless, comparison was made between 10 m*M*, 50 m*M*, and 200 m*M* AmAc buffer solutions, all containing VEGF in its native conformation. Although the samples showed a similar resolution and signal intensity, the 200 m*M* sample showed better signal stability, and thus this concentration was chosen for further work.

NanoESI TOF mass spectra of the five samples are shown in Supporting Information Figure S1. The intensities of peaks corresponding to PL or PL_2 species decreased in concordance with a decrease in K_Ds of the respective ligands. However, in all the cases it was possible to detect the peak, corresponding to the $[PL_2]^{+10}$ ion, with a *m/z* value around 2821 (here and later in the text, P, protein; PL, complex of protein with one ligand; PL_2, complex of protein with two ligands). Thus, this peak was chosen for the study of gas-phase affinity of the ligands to VEGF.

For this purpose, CID MSMS experiments were performed for each complex. The $[PL_2]^{+10}$ ions were isolated in the quadrupole section of mass-spectrometer and then transferred to the collision cell, which was filled with argon under precisely controlled pressure. Voltage bias, applied between the quadrupole and inlet electrode of collision cell (later referred to as collision voltage or Vc), was gradually increased, in order to augment the kinetic energy of the ions and provide them with internal energy, sufficient to disrupt the non-covalent bonds. The evolution of the MSMS spectrum of the VEGF:::**P-18r** complex with increase of Vc is shown in Figure 4.

Dissociation of the 1:2 VEGF:ligand complex occurred in two steps. The Vc increased from its initial value to about 40 V. This caused the parent ion to gradually dissociate, emitting one ligand. The main pathway involved dissociation of $[PL_2]^{+10}$ into $[PL]^{+8}$ and $[L]^{+2}$, and the secondary pathway—into $[PL]^{+9}$ and $[L]^{+1}$. In the case of VEGF:::**P-18r** complex dissociation of the $[PL_2]^{+10}$ ion was complete when the voltage is around 44V. Upon further increase of Vc, it starts to be enough energy for two dissociation events, and $[PL]^{+9}$ and $[PL]^{+8}$ ions started to dissociate, releasing ligand and VEGF. The Vc value of about 58 V corresponded to the total dissociation of the complex, and after this value only $[P]^{+8}$, $[P]^{+7}$, and $[P]^{+6}$ ion peaks were present in the spectrum. The most abundant one was $[P]^{+6}$, corresponding to the major dissociation pathway:

$$[PL_2]^{+10} \rightarrow [PL]^{+8} + [L]^{+2} \rightarrow [P]^{+6} + 2[L]^{+2}$$

Table II Left: Gibbs Energy and Dissociation Constants for the Ligands in Solution. Right: Collision Voltage and Internal Energy Increase for the Ligands in Gas Phase. Ligands are Arranged by the Increase in Stability

Solution			Gas Phase		
Ligand	K_D, μM	$-\Delta G$, kJ/mol	Ligand	Vc_0, V	ΔU, eV
P-10m	1810 ± 147	16.7 ± 5.1	**P-wt**	22.0 ± 1.0	64.1 ± 3.8
P-16f	313 ± 53	19.9 ± 3.4	**P-7i**	24.0 ± 1.0	68.1 ± 4.5
P-7i	252 ± 77	21.3 ± 6.2	**P-18r**	25.5 ± 1.0	75.1 ± 4.1
P-18r	3.50 ± 0.60	33.0 ± 2.5	**P-16f**	29.0 ± 1.0	81.0 ± 4.0
P-wt	1.02 ± 0.18	36.4 ± 13.0	**P-10m**	34.0 ± 1.0	94.3 ± 5.2

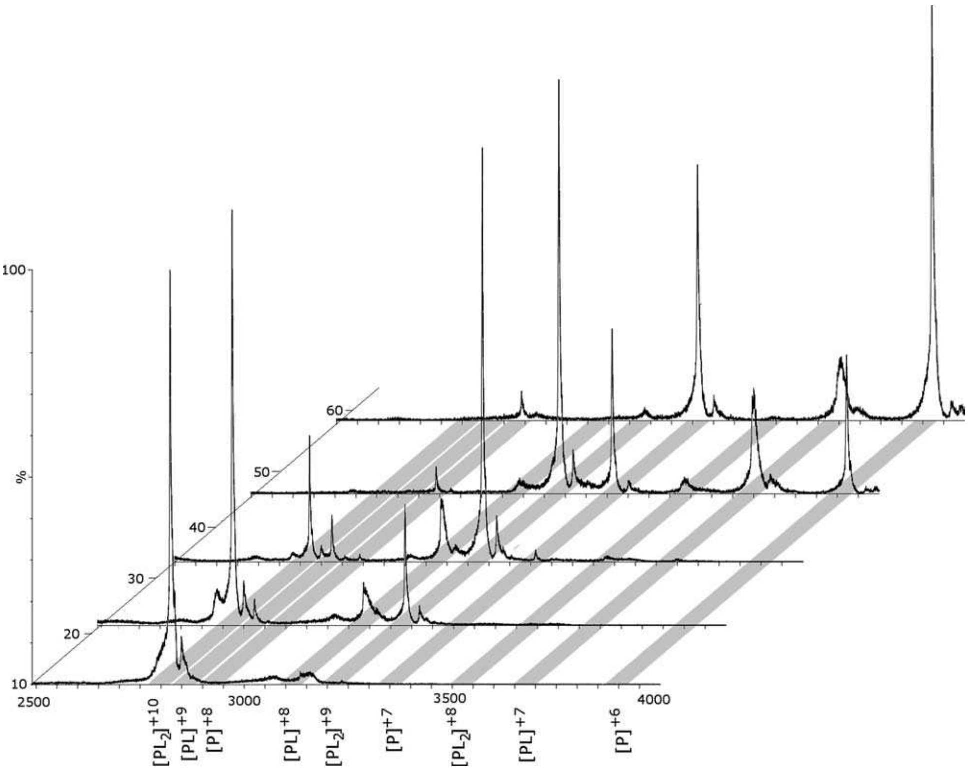


FIGURE 4 Dissociation of $[PL_2]^{+10}$ ion of VEGF:::**P-18r**, caused by increase in collision voltage from 10 to 60 V. The same peaks are highlighted with a gray stripe.

One exception was a $[PL]^{+7}$ peak, which was still present in the spectrum even under higher ion energies (Vc > 60 V). Presumably, this ion came from the minor pathway:

$$[PL_2]^{+10} \rightarrow [PL_2]^{+9} + H^+ \rightarrow [PL]^{+7} + [L]^{+2}$$

In this case, the system underwent three dissociation events to reach a fully dissociated state. Thus, more energy was required to detach both ligands from the protein.

Complexes of VEGF with each of the five ligands were examined this way, and a dissociation curve was made for each one (see Figure 5).

Internal Energy Calculation

The Impulsive Collision Transfer (ICT) model[45,46] was used to calculate the internal energy (ΔU) required for the detachment of the ligand from the complex ion in the gas phase. According to this model, it is assumed that the collision of a large macromolecule ion in the gas phase with a neutral molecule of the buffer gas involves two separate events: first, elastic (or partially inelastic) collision of a buffer gas molecule with an atom on the surface of the protein, and, second, a fully inelastic interaction of the atom of the ion that underwent collision with the rest of the molecule. The second step results in the transformation of the kinetic energy of this atom into internal energy of the ion and distribution of this energy among its internal degrees of freedom.

Here we used low-energy CID, where the ion undergoes multiple collisions before dissociation. This method belongs to the family of so-called slow-heating, or nonergodic methods,[40,41] for which it is assumed that the time, required for the energy acquired in the collision to be distributed among internal degrees of freedom is negligibly small in comparison with the time needed for the dissociation of the complex. In the ICT transfer model, the energy is distributed inside the ion immediately after the second step. For more details see the Materials and Methods section.

Ion internal energies, calculated with the ICT model, are listed in Table II. It is worth mentioning, that, strictly speaking, these energies are not "collisional" energies, that is to say they are not the energies of bonds, connecting the protein and the ligand. In fact, it is the amount of energy, transferred to the ion and distributed among its internal degrees of freedom, which appeared to be sufficient for disrupting these bonds. The question as to what proportion of this energy was used for the dissociation is a subject for further study.

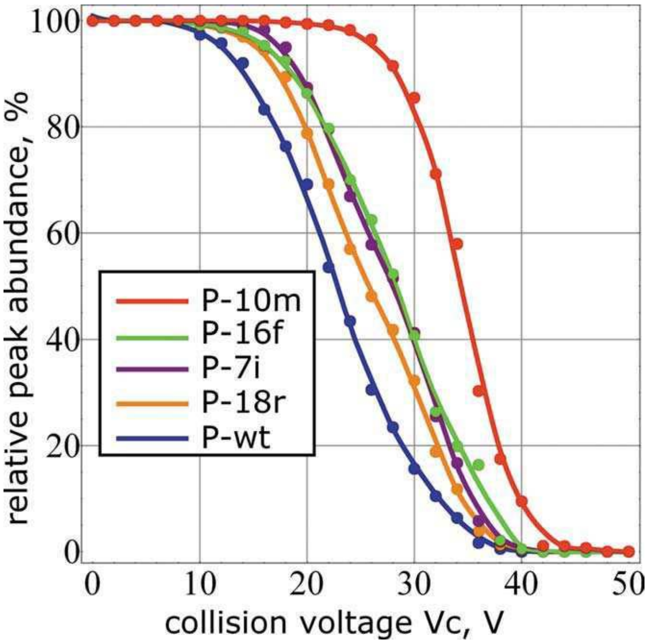


FIGURE 5 Dissociation plots of VEGF in complex with each of 5 ligands built for $[PL_2]^{+10}$ ions. Increase in collisional voltage corresponds to the increase in gas-phase binding energy and is opposite to the increase of solution binding affinity.

DISCUSSION

Protein-Ligand Complex in Solution

Table II summarizes results obtained for solution and gas-phase stabilities. In solution, when the protein functions under native conditions, the strongest ligand is **P-wt**, i.e. the "wild type" cyclopeptide, optimized for the interaction with the same construct of VEGF.[23] Although free **P-wt** is poorly ordered, when bound to VEGF it adopts a conformation with a type I β-turn and a C′-terminal α-helix. In particular, the α-helix forms part of the binding interface, pointing hydrophobic residues Trp13, Phe16, Leu19 towards the protein hydrophobic binding pocket. Together with other hydrophobic residues, buried in the binding interface, (solvent accessible surface (SAS) < 60, calculated with GETAREA web service[47]), namely, Ile7, Ala8, Met10, and Trp11, they form a highly hydrophobic interface. Inversion of configuration of any of these residues would affect the hydrophobicity of this interface, thereby decreasing its affinity to VEGF.

However, the correlation between the SAS of a given residue and the binding affinity of respective mutant ligand is not straightforward because of the effect of the change of chirality on the secondary structure of the ligand. For example, the Phe16 residue showed the lowest SAS value (<1%). However, **P-16f** exhibited a smaller decrease in binding affinity than **P-10m** (SAS 50%). This observation indicates that the change in chirality of one amino acid induces strong alterations in the conformation the ligand adopts upon binding to the protein. This can produce entropic penalty, thereby also reducing the affinity of the ligand to VEGF.

Thus, the decrease in binding affinity caused by the change in chirality of one amino acid can be attributed to both the perturbation of the hydrophobic surface and entropic penalty, caused by the loss of the pre-organized structure which may be present in the case of the wild-type ligand.

In order to gain a better insight into the nature of these changes, two ligands, namely, **P-wt** and **P-18r**, were studied by Isothermal titration calorimetry (ITC). These were the only ligands for which ITC experiments were possible, both because of insufficient solubility of ligands (free ligands in aqueous buffer at near physiological pH start to precipitate at 0.5–0.8 mM concentrations) and a large amount of protein required. VEGF (10 μM) was titrated with 170 μM **P-wt** and 247 μM **P-18r** at 37°C. Analysis of the entropy/enthalpy balance of VEGF-ligand interactions showed that although Gibbs energies (ΔG) of the two ligands differed by only 3.2 kJ/mol, differences in entropic ($T\Delta S$) and enthalpic (ΔH) effects were much stronger: 20.2 kJ/mol and 17 kJ/mol respectively (see Figure 6). The K_D values, calculated with ITC, were in good agreement with ones from NMR CSP. However, it remains unclear, why the driving force of interactions in both cases is enthalpy, whereas in case of hydrophobic interactions the main binding force should come from the entropic part.[48,49]

To address this question, the structure of the VEGF:::**P-wt** complex was examined with Molecular Operating Environment (MOE) software.[50] This program allows the prediction of noncovalent interactions that may take place between the protein and the ligand, on the basis of their relative position. For the VEGF:::**P-wt** binding interface in addition to the numerous hydrophobic interactions it predicted four hydrogen bonds: Phe16$_{\text{P-wt}}$-Gln89$_{\text{VEGF1}}$, Trp11$_{\text{P-wt}}$-Tyr21$_{\text{VEGF2}}$, Ile7$_{\text{P-}}$

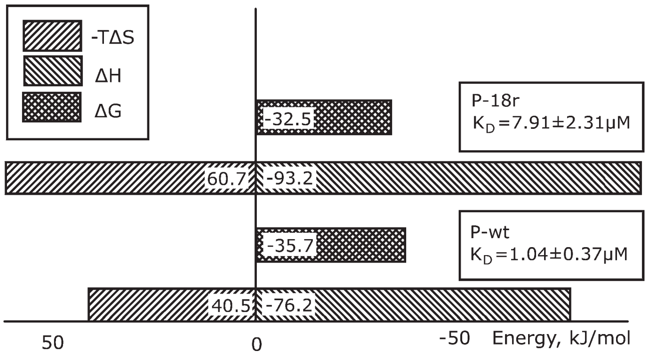


FIGURE 6 Entropy/enthalpy balance chart for interactions of VEGF with **P-wt** and **P-18r** peptides and their dissociation constants obtained with ITC.

$_{\text{wt}}$-Asn62$_{\text{VEGF2}}$, and Arg9$_{\text{P-wt}}$-Asn62$_{\text{VEGF2}}$ (lower indexes P-wt, VEGF1 and VEGF2 indicate, that the amino acid belongs to **P-wt**, 1st or 2nd monomer of VEGF dimer respectively). Thus, VEGF:::**P-wt** interactions are probably not purely hydrophobic, but also have an electrostatic component. This would explain the strong "favorable" ΔH component of ΔG.

The affinities of the two ligands were very close, although their components differed substantially. This is an example of the well-known entropy-enthalpy compensation effect.[51,52] One interpretation of the underlying processes could be the following. The perturbation of hydrophobic surface of the ligand, caused by substitution of L-Arg18 with D-Arg18, decreases the efficiency of hydrophobic interactions. The entropic component of ΔG includes the penalty for the loss of conformational freedom and "favorable" component that comes from the release of water molecules upon burial of the hydrophobic part of the ligand in the hydrophobic pocket. In **P-18r**, the latter is reduced, thereby increasing the total "unfavorable" entropic component of Δ*G*. On the other hand, perturbation of hydrophobic surface, may give more flexibility and freedom for the ligand to adopt a conformation, more favorable for hydrogen bonding. This would increase the enthalpic component of Δ*G*, which compensates the "unfavorable" entropic component.

Although the ITC was available for only two ligands, MS experiments (described in the next section) suggest that these explanations can cover the five ligands studied.

Protein-Ligand Complex in Gas Phase

The stabilities of complexes of VEGF with five ligands in gas phase, measured by CID MS, are listed in Table II. They are ranked almost exactly in the opposite way to solution values. The only exception is the **P-7i**/**P-18r** pair, but their gas-phase stabilities are relatively close. Thus, in this system, ranking of affinities in the gas phase can be considered the reverse of that in solution. The question to be addressed in this context is the meaning of the values, listed in Table II. Both Vc and ΔU are related to the total energy acquired by the ion in the collision cell before dissociation and transferred to its internal energy. However, since this energy is distributed among all degrees of freedom, it is unclear which part is devoted to breaking the bonds that keep the ligand attached.

This question is to be addressed quantitatively in the future studies. However, from the qualitative point of view, the fraction of internal energy that can be used to disrupt the protein-ligand interaction should be related to the area occupied by the ligand on the protein surface. A relatively straightforward way to check whether all the ligands bind in the same way and, thus occupy the same area in the gas phase, is to compare the collision cross-sections of the complexes. Given that all the ligands have exactly the same mass, the reliability of this experiment is increased.

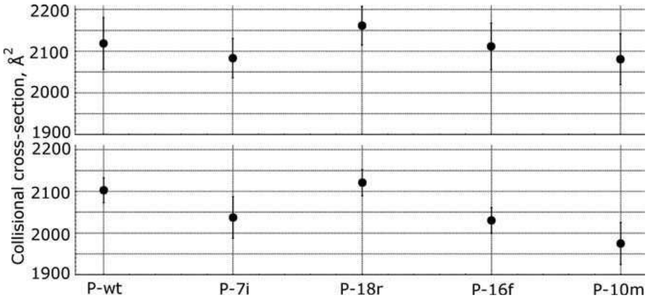


FIGURE 7 Collision cross-sections of PL_2^{+10} (bottom) and PL_2^{+11} (top) ions of complexes of VEGF with five ligands. The order from left to right corresponds to the increase in dissociation energy in gas phase.

For this experiment, an Ion Mobility Spectrometer (IMS), embedded in the Waters SYNAPT HDMS instrument, was used. The collision cross-section (Ω) of each complex was measured following the procedure described in details by Ruotulo et al.[53] As standard proteins for building the calibration curve we used equine myoglobin, equine cytochrome *c*, bovine ubiquitin and bradykinin. In total, 27 points were plotted for each set of conditions. The Ω values were measured for two sets of ions, PL_2^{+10} ($m/z = 2821$) and PL_2^{+11} ($m/z = 2564$), in nitrogen in low vacuum conditions. Although argon was used for the CID experiments, the theoretical difference between Ω in nitrogen (Ω_{N2}) and Ω in argon (Ω_{Ar}), given by Eq. (3) is less than 1%, which does not exceed the experimental error.[54] Thus, the values obtained can also be used for the calculation of ΔU (see Materials and Methods).

$$\Omega_{\mathrm{Ar}} = \Omega_{N_2}\left(\frac{R + r_{\mathrm{Ar}}}{R + r_{N_2}}\right)^2 \tag{3}$$

where r_{Ar} is the radius of argon atom, and r_{N_2} is the average "radius" of N_2 molecule. Ω values for the said complexes are plotted in Figure 7. The differences between the Ω values of the complexes are in the same order of magnitude as the experimental error, thereby making it possible to consider them similar in size. This allows us to assume, that all the ligands occupied a similar surface area on the protein, and thus, ΔU values, obtained in CID MSMS experiments, can be used to rank the ligands in terms of their capacity to remain attached to VEGF in the gas phase.

Why does this change in chirality of one amino acid cause the opposite effects on solution and gas-phase stabilities of non-covalent interactions? The main difference between the system surrounded by the solvent and one transferred to the gas phase is that in the gas phase the effects caused by the solvent are removed. In particular, hydrophobic interactions that stabilize complexes in solution are either drastically weakened or completely eliminated in the gas phase.[55] In contrast, electrostatics-based interactions, including hydrogen bonds,[56] are strengthened by the removal of solvent because of the decrease in the dielectric permittivity of the media.[36] Applying these considerations to the present case, we propose the following explanation. Perturbation of hydrophobic surface, caused by the change of chirality of one amino acid, decreases the efficiency of hydrophobic interactions. Thus, these interactions do not compensate the entropic penalty that the system pays for the loss of conformational freedom of the ligand, and the affinity of ligand in solution decreases. However, in the gas phase electrostatics plays the major role and perturbation of the hydrophobic surface may open up more possibilities for the ligand to make electrostatic contact with the VEGF binding site. Thus, the stronger the perturbation of the hydrophobic surface of the ligand, the weaker it binds to VEGF in solution, but the stronger the binding in the gas phase. This mechanism is confirmed by the ITC study, where the enthalpic part, coming from electrostatic interactions and hydrogen bonds, increases after inversion of chirality of Arg18.

The present example shows, how the CID MSMS technique provides a deeper insight into the nature of hydrophobic interactions by allowing examination of the system, unaffected by the solvent. However, we show that the assumption of similarity in behavior of non-covalent interactions in solutions and a gas phase can be made only for limited range of cases. Generally, it can be concluded, that when noncovalent interactions contain a significant hydrophobic component, gas phase affinities cannot be taken as an estimation of solution behavior either from quantitative, or qualitative point of view.

In summary, a comparison of binding in gas phase and solution can pave the way towards a better understanding of molecular recognition at protein surfaces. Peptide ligands are highly suited for these types of studies. The comparison of the relative affinities of collections of ligands with the same molecular weight facilitates the interpretation of data from MS because it is not necessary to make corrections that are not very reliable with the present state of the technique.

MATERIALS AND METHODS

Ligand Synthesis

Manual ligand synthesis was performed using a standard solid state peptide synthesis (SSPS) protocol, with Fmoc-protected amino acids (Iris Biotech GmbH, Gaurstrasse, d-95615, Marktredwitz, Germany). Other reagents, if not otherwise stated, were purchased from Carlo Erba Reagenti SpA, Strada Rivoltana, I-20090, Rodano. Dimethylformamide (DMF) and dichloromethane (DCM) solvents were used; piperidine/DMF 20%/80% (v/v) was used as a deprotection agent, TBTU (Iris Biotech) as a coupling activator, and N,N-diisopropylethylamine (DIEA, Fluka Chemie GmbH, CH-9471, Buchs, Switzerland) as an activator base. Fmoc-Rink-Amide MBHA (Iris Biotech) resin was used as a support. Cleavage from the resin was performed with trifluoroacetic acid/water/1,2-ethanedithiol/triisopropylsilane 94%/2.5%/2.5%/1%, resulting in peptides having a carboxamide group at the C-termini.

For the automatic synthesis of ligand CEM Liberty Automated Microwave peptide Synthesizer (CEM Corporation, 3100 Smith Farm Rd, Matthews, NC) was used, following standard procedures. As a support, Aminomethyl ChemMatrix resin (Matrix Innovation, 1450 City Councillors Suite 230, Montreal (Quebec), Canada), previously functionalized by Fmoc-Rink-Amide Linker (Iris Biotech) was used. As a decoupling agent, piperidine/DMF 20%/80% (v/v) with 1 m*M* of N-hydroxybenzotriazole (HOBt) was used, in order to prevent the racemization of Cys.[57] All other reagents were the same as in the manual SSPS.

Oxidation of Cys and formation of the Cys5-Cys15 disulfide bond were performed using Potassium Ferricyanide $K_3Fe(CN)_6$ (Sigma-Aldrich Co Ltd, Gillingham, Dorset, SP8 4XT, UK). Each peptide was dissolved in 5 m*M* NH_4HCO_3 aqueous buffer in a concentration of $\sim$20 mg/ml, resulting in 2 l of solution. Totally, 1 ml of 20 m*M* $K_3Fe(CN)_6$ was added every 30 min for 2 h. The reaction was monitored by Matrix-Assisted Laser Desorption Ionization (MALDI) TOF MS. After 2 h, excess solvent was removed with a rotary evaporator and the sample was freeze-dried.

MALDI TOF MS experiments were performed on an Applied Biosystems 4700 Proteomics Analyzer instrument (AB, 850 Lincoln Centre Drive Foster City, CA 94404). For the data analysis Data Explorer 4.5 was used. Each ligand was dissolved in 25/75 CH_3CN/water, purified using Waters HPLC semi preparative instrument, equipped with Waters Symmetry C18 5 μM 30 × 100 mm preparative column and freeze dried.

VEGF Expression

$VEGF_{121}{}^{11-109}$ was produced in a similar way to that described by Fairbrother et al.[29] Isotopic labeling of (methyl ^{13}C)-Met and all-^{15}N was performed by using *E. coli* strain B834 (DE3) (auxothropic for Met, purchased from Cambridge Isotope Laboratories, Inc.). Expression was carried out in M9 minimal media supplemented with 80 mg/l Met(^{13}C)and 1 g/l $^{15}NH_4Cl$.

CID MS

The CID MS experiments were carried out using a Waters SYNAPT HDMS TOF mass spectrometer with an embedded Ion Mobility Spectrometry (IMS) cell. It was equipped with TriVersa NanoMate ion source from Advion (19 Brown Road, Ithaca, NY, 14850). The instrument was operated in TOF mode. Conditions were optimized in order to maintain the noncovalent interactions in the gas phase intact.[42]

Since the mass of ligands is below the extraction limit of most of commercially available desalting columns, VEGF and a ligand for each sample were prepared separately. For the buffer exchange/desalting of the protein BioRad BioSpin-6 columns were used two times. Each ligand, previously purified using preparative HPLC with

C_{18} reversed phase column and freeze-dried, was dissolved in 200 m*M* aqueous ammonium acetate (AmAc) buffer. In order to remove traces of K^+, Millipore Microcon YM-3 centrifugal ultrafiltration devices were used. The complex solutions were prepared, with a final concentration of VEGF being 15 μM, and a ligand—60 μM for **P-wt** and **P-18r**, 90 μM for **P-7i** and **P-16f**, and 150 μM for **P-10m**. A greater concentration of weaker ligands in solution was needed to provide the sufficient abundance of PL_2 species in the sample.

Totally, 10 μl of each sample was analyzed. The pressure in the collision cell was kept equal to 3.77×10^{-2} mbar. Vc was varied from 10 V to 60 V, with 2 V increment. Hundred scans were recorded for each voltage. Spectra were processed and integrated using MassLynx v.4.1 software.

Internal Energy Calculation

The dissociation curve, i.e. the relative intensity of $[PL_2]^{+10}$ ion signal, plotted against Vc, was built for each complex. According to the ICT theory, introduced in[45,46] and applied in,[58] the increase in internal energy of the ion after single collision with a buffer gas molecule ΔU_S is given by

$$\Delta U_S = E_0 \frac{2m_g^2 m_a}{\left(M + m_g\right)\left(m_a + m_g\right)^2} \cos^2 \varphi \tag{4}$$

where E_0 is the initial kinetic energy of the ion, m_g is a mass of buffer gas molecule, m_a is a mass of colliding atom in the surface of the protein, M is the mass of the protein and φ is the angle between the trajectory of the ion and a straight line, drawn through the centers of the masses of the ion and a buffer gas molecule. After averaging, and taking into account that an ion undertakes many collisions before dissociation, total internal energy, acquired by the ion, ΔU, is given by

$$\Delta U = E_0 \left(1 - \left(1 - \frac{2m_g^2 m_a}{\left(M + m_g\right)\left(m_a + m_g\right)^2}\right)^N\right) \tag{5}$$

where N is the number of collisions, which can be calculated from the pressure in cell p and collision cross-section σ using

$$N = \frac{\sigma n l}{kT} \tag{6}$$

where l is the path of the ion before dissociation, k is a Boltzmann constant and T is temperature in Kelvin.

The E_0 was calculated from Vc_0, the collision voltage, at which the derivative of dissociation curve has a minimum. At this point, the length of the collisional cell, 0.1 m, could be taken as an l value. The calculated error is a sum of errors of cross-section measurement and Vc_0 determination.

Cross-Section Determination

Cross-sections of $[PL_2]^{+10}$ and $[PL_2]^{+11}$ complex ions for five complexes were determined as described in,[53] using a Waters SYNAPT HDMS TOF mass spectrometer, working in the ion mobility mode. N_2 buffer gas was used. For building the calibration curves, standard proteins with previously published collision cross-sections,[59] Bradykinin, Cytochrome c, Equine myoglobine, Bovine ubiquitine (Sigma Aldrich, 3050 Spruce Street, St. Louis, MO 63103) were used. Three curves were built, using three wave height values: 8.5 V, 9.5 V, and 10 V. A wave velocity of 300 m/s was used for all the cases. Data were analyzed with MassLynx 4.1 and DriftScope 2.0 software. Drift times of analyzed $[PL_2]^{+10}$ and $[PL_2]^{+11}$ ions were fitted to the resulting curves. The error of measurements was calculated, according to.[53]

K_D Determination by NMR

The NMR CSP experiments were carried out on a Bruker Digital Avance 600MHz apparatus equipped with a cryoprobe. (Bruker Daltonics Inc. 40 Manning Road, Manning Park, Billerica, MA 01821) at the High Field NMR Unit of the University of Barcelona. Samples were prepared in a 25 m*M* PB buffer containing 50 m*M* NaCl and 0.02% NaN_3 (H_2O:D_2O 9:1). The protein concentration in all cases was 100 μM. Titration with the ligands was performed by dissolving previously defined amounts of freeze-dried peptide in the sample and adding 0.5% of DMSO-d_6 to ensure solubility. ^{15}N-^{1}H HSQC spectra were acquired at 318 K using the fhsqcf3gpph pulse sequence with 2048 × 256 complex points, with a total of 8 transients per increment. The f2 domain of the dataset was increased to 512 by linear prediction and then zero-filled to yield a 2048 × 1024 data set. A qsine function was used for line broadening. All spectra were processed using Topspin 2.0.

For the fitting the resulting $\Delta\delta$ vs $[L_0]$ curves it was assumed, that change in chemical shift is proportional to the fraction of ligand, bound to the protein:

$$\Delta\delta = k\frac{[L_b]}{[P_0]} = k\frac{[L_0] - [L]}{[P_0]} \tag{7}$$

where $[L_b]$ is the concentration of ligand bound to the protein. Combining Eqs. (2) and (7) gives the final equation for the fitting:

$$\Delta\delta = k\frac{K_D + [L_0] + 2[P_0] - \sqrt{(K_D + [L_0])^2 + 4[P_0](K_D - [L_0] + [P_0])}}{2}$$

For curve fitting, a nonlinear regression procedure, built into the Wolfram Mathematica software package, was used.[60]

Isothermal Titration Calorimetry

For the ITC experiments, a VP-ITC Isothermal Titration Calorimeter was used (Microcal, 800, Centennial Avenue, Piscataway, NJ 08554) Both the ligand and the protein were dissolved in 25 m*M* PB buffer containing 50 m*M* NaCl and 0.02% NaN_3. Protein was titrated with ligand at 37°C. Volume of one injection was 10 μl, reference power, 15 kcal/s. Results were analyzed with Origin 7 script, provided by Microcal.

REFERENCES

1. Gordo, S.; Giralt, E. Protein Sci 2009, 18, 481–493.
2. Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamodar, G.; Yang, M.; Johnston, M.;

Fields, S.; Rothberg, J. M. Nature 2000, 403, 623–627.

3. Palmer, T. D.; Schwartz, P. H.; Taupin, P.; Kaspar, B.; Stein, S. A.; Gage, F. H. Nature 2001, 411, 42–43.
4. Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, Y. L.; Ooi, C. E.; Godwin, B.; Vitols, E.; Vijayadamodar, G.; Pochart, P.; Machineni, H.; Welsh, M.; Kong, Y.; Zerhusen, B.; Malcolm, R.; Varrone, Z.; Collis, A.; Minto, M.; Burgess, S.; McDaniel, L.; Stimpson, E.; Spriggs, F.; Williams, J.; Neurath, K.; Ioime, N.; Agee, M.; Voss, E.; Furtak, K.; Renzulli, R.; Aanensen, N.; Carrolla, S.; Bickelhaupt, E.; Lazovatsky, Y.; DaSilva, A.; Zhong, J.; Stanyon, C. A.; Finley, R. L., Jr.; White, K. P.; Braverman, M.; Jarvie, T.; Gold, S.; Leach, M.; Knight, J.; Shimkets, R. A.; McKenna, M. P.; Chant, J.; Rothberg, J. M. Science 2003, 302, 1727–1736.
5. Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. FEBS J 272: 5129–5148, 2005.
6. Yu, H.; Braun, P.; Yildirim, M. A.; Lemmens, I.; Venkatesan, K.; Sahalie, J.; Hirozane-Kishikawa, T.; Gebreab, F.; Li, N.; Simonis, N.; Hao, T.; Rual, J. F.; Dricot, A.; Vazquez, A.; Murray, R. R.; Simon, C.; Tardivo, L.; Tam, S.; Svrzikapa, N.; Fan, C.; de Smet, A. S.; Motyl, A.; Hudson, M. E.; Park, J.; Xin, X.; Cusick, M. E.; Moore, T.; Boone, C.; Snyder, M.; Roth, F. P.; Barabasi, A. L.; Tavernier, J.; Hill, D. E.; Vidal, M. Science 2008, 322, 104–110.
7. Carulla, N.; Caddy, G. L.; Hall, D. R.; Zurdo, J.; Gairi, M.; Feliz, M.; Giralt, E.; Robinson, C. V.; Dobson, C. M. Nature 2005, 436, 554–558.
8. Chiti, F.; Dobson, C. M. Annu Rev Biochem 2006, 75, 333–366.
9. Carulla, N.; Zhou, M.; Arimon, M.; Gairi, M.; Giralt, E.; Robinson, C. V.; Dobson, C. M. Proc Natl Acad Sci USA 2009, 106, 7828–7833.
10. Gordo, S.; Martos, V.; Santos, E.; Menendez, M.; Bo, C.; Giralt, E.; de Mendoza, J. Proc Natl Acad Sci USA 2008, 105, 16426–16431.
11. Gordo, S.; Martos, V.; Vilaseca, M.; Menéndez, M.; Mendoza, J.; Giralt, E. Chem Eur J (submitted) (2010).
12. Cruz, M.; Tusell, J. M.; Grillo-Bosch, D.; Albericio, F.; Serratosa, J.; Rabanal, F.; Giralt, E. J Pept Res 2004, 63, 324–328.
13. Grillo-Bosch, D.; Carulla, N.; Cruz, M.; Sanchez, L.; Pujol-Pina, R.; Madurga, S.; Rabanal, F.; Giralt, E. ChemMedChem 2009, 4, 1488–1494.
14. Eyleen Araya, I. O.; Neus G. B.; Simón G.; Víctor F. P.; Ernest G.; Marcelo J. K. Nanoscale Res Lett 2008, 3, 435–443.
15. Olmedo, I.; Araya, E.; Sanz, F.; Medina, E.; Arbiol, J.; Toledo, P.; Alvarez-Lueje, A.; Giralt, E.; Kogan, M. J. Bioconjug Chem 2008, 19, 1154–1163.
16. Kogan, M. J.; Bastus, N. G.; Amigo, R.; Grillo-Bosch, D.; Araya, E.; Turiel, A.; Labarta, A.; Giralt, E.; Puntes, V. F. Nano Lett 2006, 6, 110–115.
17. Dudley, H. W.; Cox J. P. L.; Andrew, J. D.; Mark, G.; Ute, G.; Perry, T. K.; Allick, R. L.; Ian, A. N.; Colin, J. S.; Robert, C. M. J Am Chem Soc 1991, 113, 7020–7030.
18. Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Angew Chem Int Ed Engl 2004, 43, 6596–6616.
19. Whitesides, G. M.; Krishnamurthy, V. M. Q Rev Biophys 2005, 38, 385–395.
20. Gilson, M. K.; Zhou, H. X. Annu Rev Biophys Biomol Struct 2007, 36, 21–42.
21. Nicolas, E.; Ferrer, C.; Taboada, L.; Giralt, E. J Am Chem Soc 2005, 127, 17719–17733.
22. Pastor, J. J.; Granados, G.; Carulla, N.; Rabanal, F.; Giralt, E. J Am Chem Soc 2007, 129, 14922–14932.
23. Pan, B.; Li, B.; Russell, S. J.; Tom, J. Y.; Cochran, A. G.; Fairbrother, W. J. J Mol Biol 2002, 316, 769–787.
24. Hoeben, A.; Landuyt, B.; Highley, M. S.; Wildiers, H.; Van Oosterom, A. T.; De Bruijn, E. A. Pharmacol Rev 2004, 56, 549–580.
25. Ferrara, N.; Kerbel, R. S. Nature 2005, 438, 967–974.
26. Ferrara, N. Am J Physiol Cell Physiol 2001, 280, C1358–C1366.
27. Lu, N.; Gao, Y.; Ling, Y.; Chen, Y.; Yang, Y.; Gu, H. Y.; Qi, Q.; Liu, W.; Wang, X. T.; You, Q. D.; Guo, Q. L. Life Sci 2008, 82, 956–963.
28. Brazelle, W. D.; Shi, W.; Siemann, D. W. Int J Radiat Oncol Biol Phys 2006, 65, 836–841.
29. Fairbrother, W. J.; Champe, M. A.; Christinger, H. W.; Keyt, B. A.; Starovasnik, M. A. Protein Sci 1997, 6, 2250–2260.
30. Pal, G.; Kossiakoff, A. A.; Sidhu, S. S. J Mol Biol 2003, 332, 195–204.
31. Bonvin, A. M.; Boelens, R.; Kaptein, R. Curr Opin Chem Biol 2005, 9, 501–508.
32. Fielding, L. Curr Top Med Chem 2003, 3, 39–53.
33. Bristow, A. W.; Webb, K. S. J Am Soc Mass Spectrom 2003, 14, 1086–1098.
34. Baldwin, M. A. Methods Enzymol 2005, 402, 3–48.
35. Siuzdak, G.; Bothner, B.; Yeager, M.; Brugidou, C.; Fauquet, C. M.; Hoey, K.; Chang, C. M. Chem Biol 1996, 3, 45–48.
36. Nesatyy, V. J. Int J Mass Spectr 2002, 221, 147–161.
37. Jürg, M. D.; S. D. F.; Rajagopalan, S.; Silke W.; Renato, Z. Int J Mass Spectrom 2002, 216, 1–27.
38. Loo, J. A. Int J Mass Spectrom 2000, 200, 175–186.
39. Zhang, S.; Van Pelt, C. K.; Wilson, D. B. Anal Chem 2003, 75, 3010–3018.
40. Wells, J. M.; McLuckey, S. A. Methods Enzymol 2005, 402, 148–185.
41. Mayer, P. M.; Poon, C. Mass Spectrom Rev 2009, 28, 608–639.
42. Hernandez, H.; Robinson, C. V. Nat Protoc 2007, 2, 715–726.
43. Iavarone, A. T.; Udekwu, O. A.; Williams, E. R. Anal Chem 2004, 76, 3944–3950.
44. Verkerk, U. H.; Kebarle, P. J Am Soc Mass Spectrom 2005, 16, 1325–1341.
45. Uggerud, E.; Derrick, P. J. J Phys Chem 1991, 95, 1430–1436.
46. Helen J. C.; P. J. D.; Donald, H.; Jenkins, B. J Phys Chem 1993, 97, 5443–5444.
47. Fraczkiewicz, R.; Braun W. J Comput Chem 1998, 19, 319–333.
48. Finkelstein, A. V.; Ptitsyn, O. B. Protein Physics: A Course of Lectures; Academic Press, San Diego, California, 2002.
49. Santos, H. A.; Manzanares, J. A.; Murtomaki, L.; Kontturi, K. Eur J Pharm Sci 2007, 32, 105–114.
50. MOE, Chemical Computing Group Inc.: Montreal, available at: http://www.chemcomp/com.
51. Dunitz, J. D. Chem Biol 1995, 2, 709–712.
52. Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nollmann, M. Biophys Chem 2001, 93, 215–230.
53. Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. Nat Protoc 2008, 3, 1139–1152.
54. Asbury, G. R.; Hill, H. H., Jr. Anal Chem 2000, 72, 580–584.
55. Yin, S.; Xie, Y.; Loo, J. A. J Am Soc Mass Spectrom 2008, 19, 1199–1208.
56. Dannenberg, J. J.; Haskamp, L.; Masunov, A. J Phys Chem 1999, 103, 7083–7086.
57. Palasek, S. A.; Cox, Z. J.; Collins, J. M. J Pept Sci 2007, 13, 143–148.
58. Serguei Ilchenko, R. J. C. Int J Mass Spectrom 2007, 265, 372–381.
59. Clemmer Cross Section Database, available at: http://www.

3.

Quantitative assessment of gas-phase stabilities of non-covalent complexes.

Quantitative assessment of gas-phase stabilities of non-covalent complexes.

(Submitted to J. Amer. Chem. Soc)

Andrey Dyachenko, Ernest Giralt

ABSTRACT

Non-covalent interactions that involve biological macromolecules are essential for vital functions of living organisms. Over recent years, considerable effort has been devoted to the development of techniques that allow the study of protein non-covalent complexes in both solution and gas phase. In contrast to a variety of solution methods, offering both qualitative and quantitative information, gas-phase strategies cannot determine the energy that stabilizes non-covalent complexes in the absence of solvent. Here we present a novel approach to determine the activation energy of gas-phase unimolecular dissociation from collision-induced dissociation mass spectrometry experiments. It combines basic physical principles with the Monte Carlo simulation technique in order to model ion excitation/relaxation upon collisions in the collision cell. The method is easily customizable and can be applied to a variety of collision cells found in modern beam-type instruments, including travelling wave ion cells. The approach forms a basis for accurate determination of gas-phase binding stabilities for non-covalent protein-protein and protein-ligand complexes that are accessible with modern mass-spectrometry instrumentation.

Biological macromolecules and metabolites perform their functions in living organisms largely through the weak non-covalent interactions between each other [1]. The task of building a human interaction network is comparable with that of the Human Genome Project in terms of its importance and complexity [2]. On the other hand, deciphering the mechanism underlying each particular biological non-covalent interaction poses a major challenge. A better understanding of the principles that define the strength, specificity and kinetics of non-covalent interactions would allow the development of more specific and efficient drugs against many diseases. Recently, mass-spectrometry made its way into biological sciences and, in particular, drug discovery, offering features such as speed, sensitivity and high throughput for screening and structure-activity relationship studies [3]. In particular, native mass-spectrometry in combination with various dissociation techniques allows the study of intact non-covalent complexes and the evaluation of their stabilities in the gas phase. It allows the assessment of affinity and stoichiometry of interactions when only nanograms of sample are available, and in-detail comparison between interactions in gas phase and solution when the sample is abundant [4–7].

There is, however, a major drawback that prevents mass spectrometry from being widely accepted as a technique of choice for drug discovery. Gas phase stabilities of non-covalent interactions cannot be assessed directly, but rather via experimental parameters, such as accelerating voltages and pressure of target gas. This limitation is because the internal energy deposited in an ion during collisional activation is difficult to determine [6], [8], [9]. Here we propose a method that allows extraction of the gas-phase stabilities of non-covalent molecular complexes from results of collision-induced dissociation (CID) experiments with no restriction for molecular weight. In its current implementation, it is used to model beam-type instruments with travelling wave technology (such as Waters

SYNAPT G1 and G2) or without it (such as Applied Biosystems Qstar). However, it can be extended to trap-type instruments.

Dissociation of a non-covalent complex upon collision can be viewed as a two-step process: first, the activation step, which involves transformation of the kinetic energy of the ion into internal energy as a result of collisions with target gas molecules, and an activated complex is formed [10]; and second, unimolecular dissociation of the activated complex. In the case of CID of large (tens to hundreds kDa) non-covalent complexes, the activation step involves slow heating of the complex during a large number of collisions [11]. This is due to the relatively low voltages generally used in native mass-spectrometry experiments and the large difference between the masses of the ion and the target gas molecules, resulting in low center of mass (COM) kinetic energy of the ion-molecule system and low energy transmission in a single collision. The dissociation step is usually modeled via the Rice–Ramsperger–Kassel–Marcus (RRKM) theory [7], which allows calculation of a dissociation rate constant provided that the necessary parameters, such as the activation energy E_A of dissociation and the density of energy states (DOS), are known. The activation step, on the other hand, is more difficult to describe, and a universal approach that can be applied to any CID experiment has not been developed to date. A number of techniques have been devised that are able to successfully describe the collisional activation of small molecules [12], [13] or small peptides [14] with subsequent breakage of covalent bonds. However, the accurate determination of the internal energy acquired by large non-covalent complexes in CID experiments still poses a major hurdle for the quantitative description of non-covalent chemistry in gas phase.

Here we present a method that combines experimental data with Monte Carlo simulations to model the deposition of internal energy during collisional activation of a molecular ion. The method consists of two parts. The first is a general collisional model, based on the probability of an increase or decrease of internal energy of a molecular ion upon collision for an arbitrary range of ion internal energies and center of mass kinetic energies. This model is thus applicable to virtually any shape of CID cell that is currently on the market, with no theoretical restriction to the molecular weight of the ion, thus making it suitable for the study of large non-covalent complexes. The second part describes spatial trajectories of ions in the collision cell. This part is highly customizable and can be adapted to various collision cell configurations found in beam-type instruments. Here we applied our method to the traveling wave ion guide (TWIG) collision cell found in the popular SYNAPT G1 (Waters Inc.) mass-spectrometer. The performance of the method was tested on CID of a small peptide (leu-enkephalin) and then applied to a series of non-covalent peptide-protein (vascular endothelial growth factor, VEGF) complexes (Figure 1A and Table 1).

Table 1*.* List of peptides used as ligands of VEGF. Lower case letter in bold indicates a **D** amino acid. Values of activation energies for corresponding complexes are shown.

Name	***Sequence***	***E_A, eV***
P-wt	GGNECDIARMWEWECFERL	**8.64 ± 0.63**
P-7i	GGNECD**i**ARMWEWECFERL	**8.31 ± 0.61**
P-18r	GGNECDIARMWEWECFE**r**L	**9.74 ± 0.71**
P-16f	GGNECDIARMWEWEC**f**ERL	**9.58 ± 0.70**
P-10m	GGNECDIAR**m**WEWECFERL	**13.73 ± 1.00**

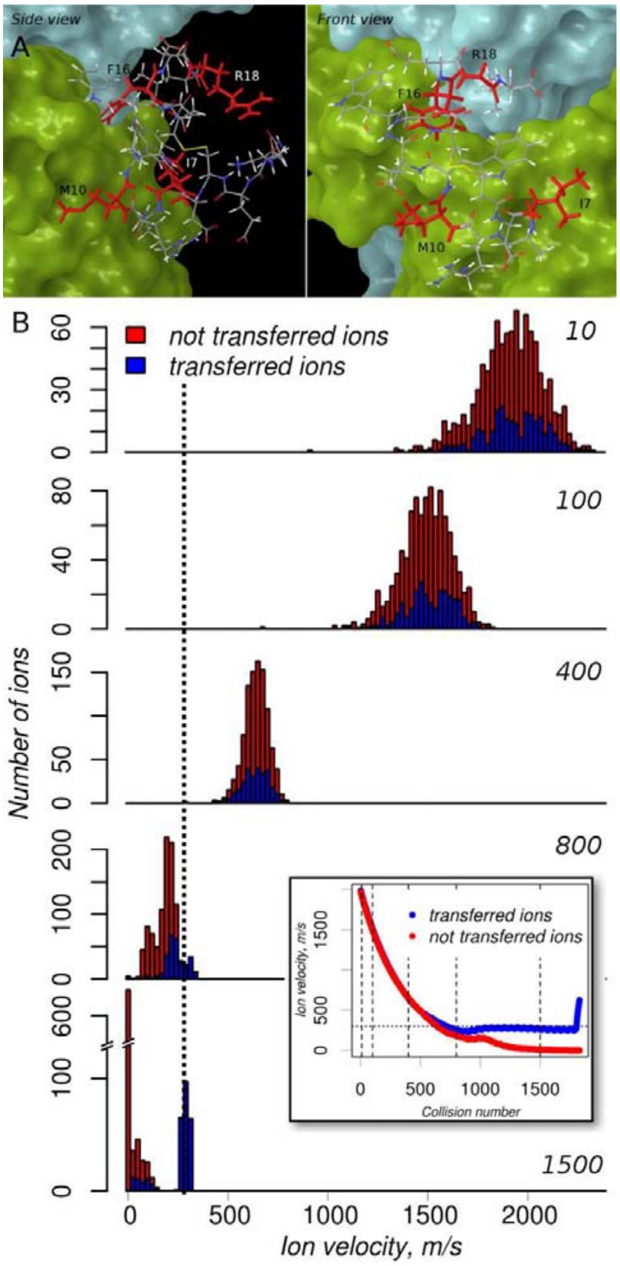


Figure 1. (A) Peptide P-wt bound to the VEGF binding site. VEGF subunits are shown in green and light blue. Modified peptide amino acids are in red. PDB entry 1KAT. (B) Velocity distribution of ions in travelling wave ion guide after a different number of collisions. Blue bars correspond to ions that successfully reached the exit of the cell and red bars ions that stopped before that. Dotted line shows the travelling wave velocity. On the inset: average ion velocity plotted against number of collisions. Dashed lines show points at which histograms are plotted.

A proper collisional model is crucial to achieve an accurate description of energy deposition during collisional activation. In the case of large non-covalent protein-protein or protein-ligand complexes, this question is further complicated by the inability, with current computational methods, to explicitly calculate DOS [15]. A number of collisional models have been proposed [16–20]; however, although some successfully describe the collisional activation of small molecules, it is difficult to adapt them to larger non-covalent complexes.

Our model is based on the approach described in [21] for the collisional activation of small molecules. Energy exchange between ion internal and translational degrees of freedom upon collision is treated as quasi steady state statistical process. For a full description of the collisional activation of an ensemble of ions, it is sufficient to know the function $P(E,K_{COM},\Delta E)$, which describes the probability of the ion with internal energy E in a collision with center of mass (COM) kinetic energy K_{COM} undergoing the energy step with size ΔE [22]. Once this function is known, it can be used as an acceptance criterion for a Monte Carlo simulation algorithm.

In order to find $P(E,K_{COM},\Delta E)$, a detailed balance principle [23] is applied and multivariate integral equation is constructed (eq. S10). We developed a novel numerical method that allows solution of the equation on a grid of (E,K_{COM}) values with arbitrary precision, defined by the step size. A Monte Carlo simulation method was devised (Figure S1A), in which $P(E,K_{COM},\Delta E)$ is used to determine for each collision whether an ion is excited, cooled or whether the collision is elastic. $P(E,K_{COM},\Delta E)$ is also used indirectly to determine the size of the positive or negative energy step. More details can be found in SI.

In the TWIG the voltage pulse produced by the array of stacked electrodes is superimposed on confining RF field [24]. Ion propulsion along the axis is achieved when the ion is “surfing” the wave, i.e. is being pushed in front of it (Figure S1B,C). In the present method, a simplified picture of TWIG is constructed that does not account for the focusing RF electric field and considers only the wave. Nevertheless, we show that this approximation, although much less accurate than simulations with dedicated ion optics modeling software like SIMION (S. I. S. Inc.,), still provides a satisfactory estimation of experimental data at a much lower computational cost. The simulation algorithm for the Waters SYNAPT G1 standard collision cell is described in details in SI.

The evolution of the velocities of 1000 ions with a mass of 28 kDa and charge +10 in TWIG, simulated at V_{COL} 30 V, wave velocity 300 m/s and wave pulse amplitude 0.5 V is shown in Figure 1B. The inset shows average velocities for ions that passed through the cell and for those that stopped before. Successful ions find resonance with the wave, as indicated by the wave-like shape of the curve. Their speed is maintained around 300 m/s, indicating that most of the ions “surf” on single wave pulse after equilibration. At the end of the cell they are accelerated by the exit electrode. In contrast, unsuccessful ions quickly lose their velocity and decelerate.

This process is further illustrated by the bar plots, showing the distribution of velocities after different numbers of collisions. Ions start with a broad distribution of velocities, which narrows as they lose their kinetic energy in collisions. At higher velocities, distributions of successful and unsuccessful ions are the same. After 800 collisions, when ions reach a steady state, the discrepancy starts to appear. The speed of successful ions is maintained closer to the wave velocity, whereas that of unsuccessful ions keeps decreasing. Finally, after 1500 collisions, the speed of successful ions is distributed mainly around the wave velocity, while that of unsuccessful ion is equal or close to zero. Ion transmission efficiency, calculated after 10 simulations containing 1000 ion trajectories each, was 28.2 ± 3.1 %. This value is similar to experimental results for large ions [25].

The experimental data were processed following the scheme in Figure 2A.

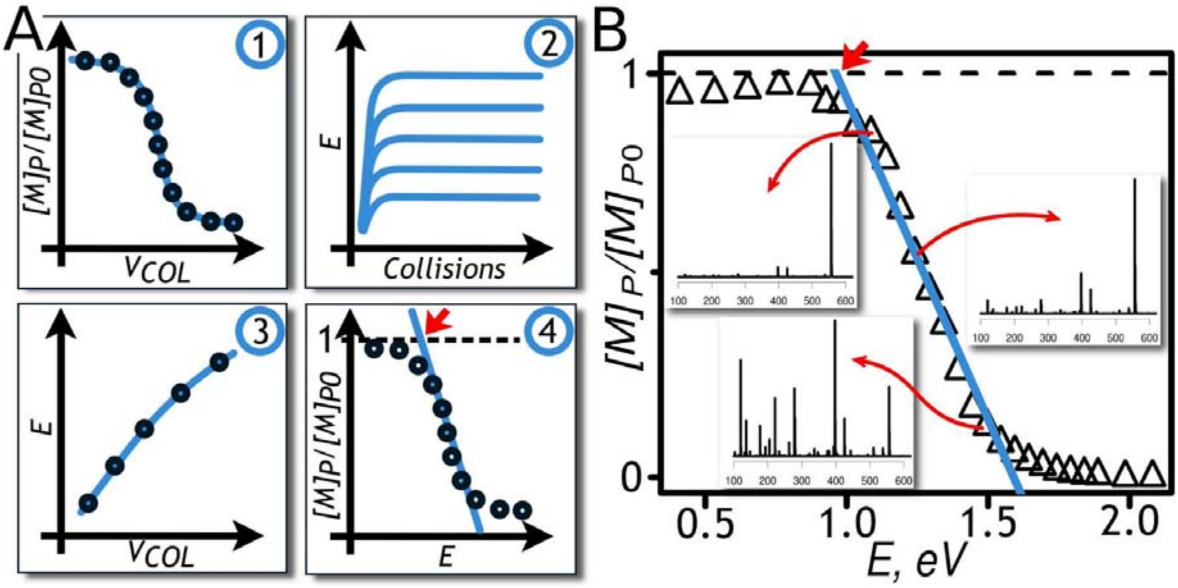


Figure 2. (A) The scheme of the experimental data treatment workflow. (B) Internal energy-dependent breakdown curve of leu-enk with spectra, indicated for three points.

A collision voltage-dependent breakdown curve [26] is constructed from the raw data following eq. S30 (1). The median value of collision voltage V_{COL} is found and used as an input parameter to optimize the collisional cooling equation and to ensure that the system maintains a steady state. After completion of the optimization, a series of simulations is performed at several V_{COL} covering the range of experimental values, and internal energy E, deposited in ions at each V_{COL}, is determined (2). From these results, the correlation between V_{COL} and E is parametrized in the form of eq. S28 (3). This allows the construction of a “real” - internal energy-dependent breakdown curve [26] (4). The activation energy of the reaction is found by extrapolating the linear region of the curve up to the intersection with $[M]_P/[M]_{P0}$ = 1. E at the point of intersection is considered the activation energy E_A. More details can be found in SI.

Validation of the model is not a trivial task because the reliable data on energy deposition in large non-covalent protein complexes available in the literature is limited [27–30]. On the other hand, the model should perform equally well for moderate sized molecules that are small enough to allow an exact computation of their density of states but still considerably larger than target gas molecules. Short peptides appear to be good candidates, and here we chose leu-enkephalin (YGGFL, molecular weight 556.28 Da), Probably the most widely studied peptide, leu-enkephalin became a standard for many mass-spectrometry techniques [31] and many experimental parameters are available for it.

Experimental V_{COL}-dependent breakdown curves were obtained for leu-enkephalin on Waters Synapt G1 instrument, working in the TOF mode (Figure S6). Spectra were recorded at V_{COL} ranging from 2 to 34 V with increments of 1 or 2 V. (Figure S5A). After optimization of the collisional cooling function, a series of simulations was performed at V_{COL} ranging from 5 to 35 V with 5 V increment, and correlation between collision voltage and internal energy E, deposited in the ion, was established. This approach allowed construction of an internal energy-dependent breakdown curve (Figure 2B). Activation energy E_A was found to be 1.01 ± 0.05 eV. This is in good agreement with published values for leu-enkephalin (1.14 ± 0.05 eV) [31]. The small discrepancy may arise from the oversimplified treatment of collision cross-section as a constant value, independent of ion kinetic energy. This error is expected to decrease for larger ions.

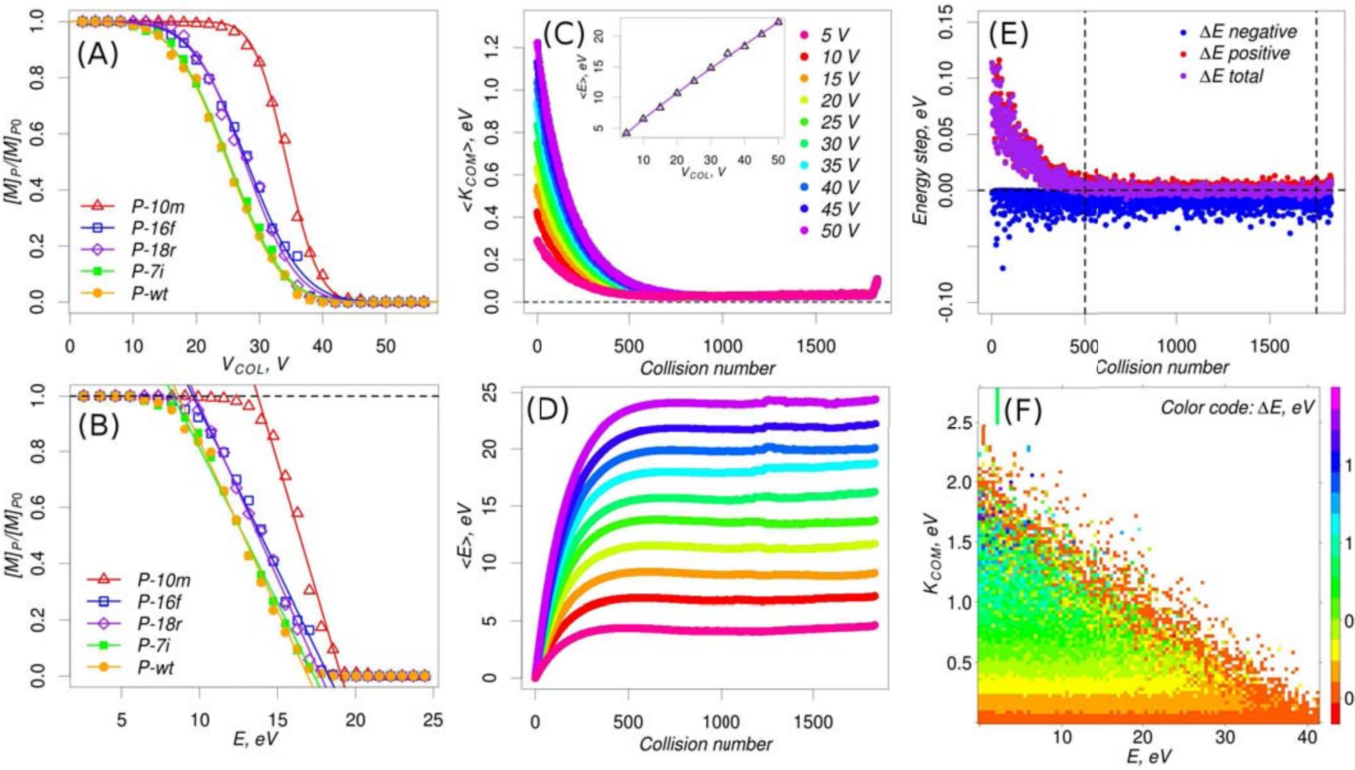


Figure 3. Treatment of 5 non-covalent complexes of VEGF with peptides. (A) Experimental collision voltage-dependent breakdown curves (points) and fitted curves (lines). (B) Internal energy-dependent breakdown curves (points) and fitting of their linear segments (lines). (C) Average COM kinetic energy of ion-molecule system, plotted against number of collisions. (D) Average internal energy of the ion, plotted against number of collisions. (E) Average positive, negative and total energy steps, plotted against number of collisions. (F) Level plot of energy steps generated during ten simulation runs and plotted on a grid of COM kinetic energies and internal energies.

Non-covalent interactions between vascular endothelial growth factor A (VEGF) and five synthetic cyclopeptides (Table 1, Figure 1A) were examined as a case study for the method. Structural details and the experimental data on the study of these interactions by CID-MS were published a few years ago [32]. VEGF is a covalent homodimer with a molecular weight of 23258.6 Da. Each VEGF molecule can bind two cyclic peptides (m.w. 2342.6 Da), forming a non-covalent complex with a molecular weight of 27943.8 Da. The pressure in the collision cell was $3.77\text{x}10^{-2}$ mbar. Spectra were recorded on a Waters SYNAPT G1 instrument at V_{COL} ranging from 8 to 44 V with 2 V increments (Figure 3A).

Parameter optimization was performed via a series of simulations, each sampling 200 trajectories. At α_{-} = 23, the system maintained a steady state with average energy change upon collision of 0.2 ± 2.3 meV (Figure 3E). Internal energy deposition was modeled at V_{COL} ranging from 5 V to 50 V with 5 V increments. Figures 3C and D show the values of internal energy and COM kinetic energy, averaged over 1000 trajectories and plotted against number of collisions, experienced by the ion. Ion kinetic energy was gradually transferred to its internal energy until the steady state was reached. It is worth noting that although the maximum energy that can be deposited in the ion upon a single collision is limited by the COM kinetic energy of the system, energy deposited in multiple collisions can far exceed COM kinetic energy, as is often the case with slow heating of large ions [9].

The average energy values $<E>$ were then correlated to corresponding V_{COL} through a power function (Figure 3C, inset), and the energy-dependent breakdown curve was constructed. A high degree of similarity between the five complexes allowed us to use the same correlation data to construct energy-dependent breakdown curves for all five. The curves are shown in Figure 3B. In order to extract non-covalent complex dissociation activation energy E_A, the linear region of each curve was found and extrapolated until intersection with the line $[M]_P/[M]_{P0}$ = 1. The activation energies for the five peptides are showSGR-n in Table 1.

In this communication we propose a method to treat CID experimental data that allows the collection of accurate information about the gas phase stability of large molecular ions. Parameters such as internal energy deposition (Figure 3F) and unimolecular dissociation activation energy can be estimated for a wide range of conditions. Although this model can be applied to molecular ions of any size, it performs particularly well with large systems, thus making it suitable to study non-covalent interactions between biomolecules in the gas phase.

ASSOCIATED CONTENT

Supporting Information

Details of mathematical model, Monte Carlo algorithm and TWIG modeling;
leu-enkephalin experimental and computational results;
Program source code and data processing scripts can be supplied upon request.

AUTHOR INFORMATION

Corresponding authors

andrey.dyachenko@irbbarcelona.org and ernest.giralt@irbbarcelona.org

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by *Generalitat de Catalunya* (XRB 2009-SGR-1005), and MCI-FEDER (BIO2008-00799). A.D. is supported by "LaCaixa/IRB Barcelona" international PhD program fellowship. We thank Tanya Yates for her editorial support.

References

[1] M. Baker, "Proteomics: The interaction map," *Nature*, pp. 2–5, 2012.

[2] Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter, T. Maniatis, A. Califano, and B. Honig, "Structure-based prediction of protein-protein interactions on a genome-wide scale.," *Nature*, vol. 490, no. 7421, pp. 556–60, Oct. 2012.

:3] M. Schirle, M. Bantscheff, and B. Kuster, “Mass spectrometry-based proteomics in preclinical drug discovery.,” *Chemistry & biology*, vol. 19, no. 1, pp. 72–84, Jan. 2012.

[4] K. J. Pacholarz, R. A. Garlish, R. J. Taylor, and P. E. Barran, “Mass spectrometry based tools to investigate protein-ligand interactions for drug discovery.,” *Chemical Society reviews*, vol. 41, no. 11, pp. 4335–55, Jun. 2012.

[5] C. Bich and R. Zenobi, “Mass spectrometry of large complexes.,” *Current opinion in structural biology*, vol. 19, no. 5, pp. 632–9, Oct. 2009.

[6] N. Morgner and C. V Robinson, “Linking structural change with functional regulation-insights from mass spectrometry.,” *Current opinion in structural biology*, vol. 22, no. 1, pp. 44–51, Feb. 2012.

[7] V. Nesatyy, “Mass spectrometry evaluation of the solution and gas-phase binding properties of noncovalent protein complexes,” *International Journal of Mass Spectrometry*, no. 221, pp. 147–161, 2002.

[8] M. Demireva and E. R. Williams, “Measuring internal energy deposition in collisional activation using hydrated ion nanocalorimetry to obtain peptide dissociation energies and entropies.,” *Journal of the American Society for Mass Spectrometry*, vol. 21, no. 7, pp. 1133–43, Jul. 2010.

[9] P. M. Mayer and C. Poon, “The mechanisms of collisional activation of ions in mass spectrometry.,” *Mass spectrometry reviews*, vol. 28, no. 4, pp. 608–39, 2008.

[10] J. L. P. Benesch, “Collisional activation of protein complexes: picking up the pieces.,” *Journal of the American Society for Mass Spectrometry*, vol. 20, no. 3, pp. 341–8, Mar. 2009.

[11] S. A. McLuckey and D. E. Goeringer, “SPECIAL FEATURE:TUTORIAL Slow Heating Methods in Tandem Mass Spectrometry,” *Journal of Mass Spectrometry*, vol. 32, no. 5, pp. 461–474, May 1997.

[12] D. Nilsson and S. Nordholm, “Modeling energy transfer in molecular collisions: Statistical theory versus experiment for highly excited toluene and azulene,” *The Journal of Chemical Physics*, vol. 119, no. 21, p. 11212, Dec. 2003.

[13] L. Drahos and K. Vékey, “MassKinetics: a theoretical model of mass spectra incorporating physical processes, reaction kinetics and mathematical descriptions.,” *Journal of mass spectrometry : JMS*, vol. 36, no. 3, pp. 237–63, Mar. 2001.

[14] O. Meroueh and W. L. Hase, “Energy transfer pathways in the collisional activation of peptides,” *International Journal of Mass Spectrometry*, vol. 201, no. 1–3, pp. 233–244, Jul. 2000.

[15] S. Singh, M. Chopra, and J. J. de Pablo, “Density of states-based molecular simulations.,” *Annual review of chemical and biomolecular engineering*, vol. 3, pp. 369–94, Jan. 2012.

[16] E. Uggerud and P. Derrick, “Theory of collisional activation of macromolecules. Impulsive collisions of organic ions,” *The Journal of Physical Chemistry*, no. 4, pp. 1430–1436, 1991.

[17] L. Börjesson and S. Nordholm, “A partially ergodic multiple encounter theory of collisional energy transfer,” *Chemical physics*, vol. 4, no. 212, pp. 393–408, 1996.

:18] L. Börjesson, L. Ming, and S. Nordholm, “The PEMET model of collisional energy transfer in unimolecular reactions comparison with molecular dynamics simulation,” *Chemical Physics*, vol. 221, pp. 253–267, 1997.

[19] D. E. Goeringer and S. A. McLuckey, “Evolution of ion internal energy during collisional excitation in the Paul ion trap: A stochastic approach,” *The Journal of Chemical Physics*, vol. 104, no. 6, p. 2214, Feb. 1996.

[20] Y. L. Chen, J. M. Campbell, B. A. Collings, L. Konermann, and D. J. Douglas, “Stability of a highly charged noncovalent complex in the gas phase: holomyoglobin.,” *Rapid communications in mass spectrometry : RCM*, vol. 12, no. 15, pp. 1003–10, Jan. 1998.

[21] W. R. Plass and R. G. Cooks, “A model for energy transfer in inelastic molecular collisions applicable at steady state or non-steady state and for an arbitrary distribution of collision energies.,” *Journal of the American Society for Mass Spectrometry*, vol. 14, no. 12, pp. 1348–59, Dec. 2003.

[22] T. Lenzer, K. Luther, K. Reihs, and A. C. Symonds, “Collisional energy transfer probabilities of highly excited molecules from kinetically controlled selective ionization (KCSI). II. The collisional relaxation of toluene: P(E[sup ′],E) and moments of energy transfer for energies up to 50000 cm−1,” *The Journal of Chemical Physics*, vol. 112, no. 9, p. 4090, Mar. 2000.

[23] A. Verbeure, “Detailed balance and equilibrium,” *Communications in Mathematical Physics*, vol. 95, no. 3, pp. 301–305, Sep. 1984.

[24] K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, and R. H. Bateman, “Applications of a travelling wave-based radio-frequency-only stacked ring ion guide.,” *Rapid communications in mass spectrometry : RCM*, vol. 18, no. 20, pp. 2401–14, Jan. 2004.

[25] K. Giles, J. L. Wildgoose, D. J. Langridge, and I. Campuzano, “A method for direct measurement of ion mobilities using a travelling wave ion guide,” *International Journal of Mass Spectrometry*, vol. 298, no. 1–3, pp. 10–16, Dec. 2010.

[26] K. Vékey, “Internal Energy Effects in Mass Spectrometry,” *Journal of Mass Spectrometry*, vol. 31, no. 5, pp. 445–463, May 1996.

[27] J. M. Daniel, S. D. Friess, S. Rajagopalan, S. Wendt, and R. Zenobi, “Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry,” *International Journal of Mass Spectrometry*, vol. 216, no. 1, pp. 1–27, Apr. 2002.

[28] P. M. Mayer and E. Martineau, “Gas-phase binding energies for non-covalent Aβ-40 peptide/small molecule complexes from CID mass spectrometry and RRKM theory.,” *Physical chemistry chemical physics : PCCP*, vol. 13, no. 11, pp. 5178–86, Mar. 2011.

[29] V. J. Nesatyy, “Gas-phase binding of non-covalent protein complexes between bovine pancreatic trypsin inhibitor and its target enzymes studied by electrospray ionization tandem mass spectrometry.,” *Journal of mass spectrometry : JMS*, vol. 36, no. 8, pp. 950–9, Aug. 2001.

[30] M. Schäfer, C. Schmuck, M. Heil, H. J. Cooper, C. L. Hendrickson, M. J. Chalmers, and A. G. Marshall, “Determination of the activation energy for unimolecular dissociation of

a non-covalent gas-phase peptide: substrate complex by infrared multiphoton dissociation fourier transform ion cyclotron resonance mass spectrometry.," *Journal of the American Society for Mass Spectrometry*, vol. 14, no. 11, pp. 1282–9, Nov. 2003.

[31] J. Sztáray, A. Memboeuf, L. Drahos, and K. Vékey, "Leucine enkephalin--a mass spectrometry standard.," *Mass spectrometry reviews*, vol. 30, no. 2, pp. 298–320, 2011.

[32] A. Dyachenko, M. Goldflam, M. Vilaseca, and E. Giralt, "Molecular recognition at protein surface in solution and gas phase: Five VEGF peptidic ligands show inverse affinity when studied by NMR and CID-MS.," *Biopolymers*, vol. 94, no. 6, pp. 689–700, Jan. 2010.

Quantitative assessment of gas-phase stabilities of non-covalent complexes.

Andrey Dyachenko, Ernest Giralt

Supporting information

Table of contents:

1. Monte-Carlo simulation algorithm

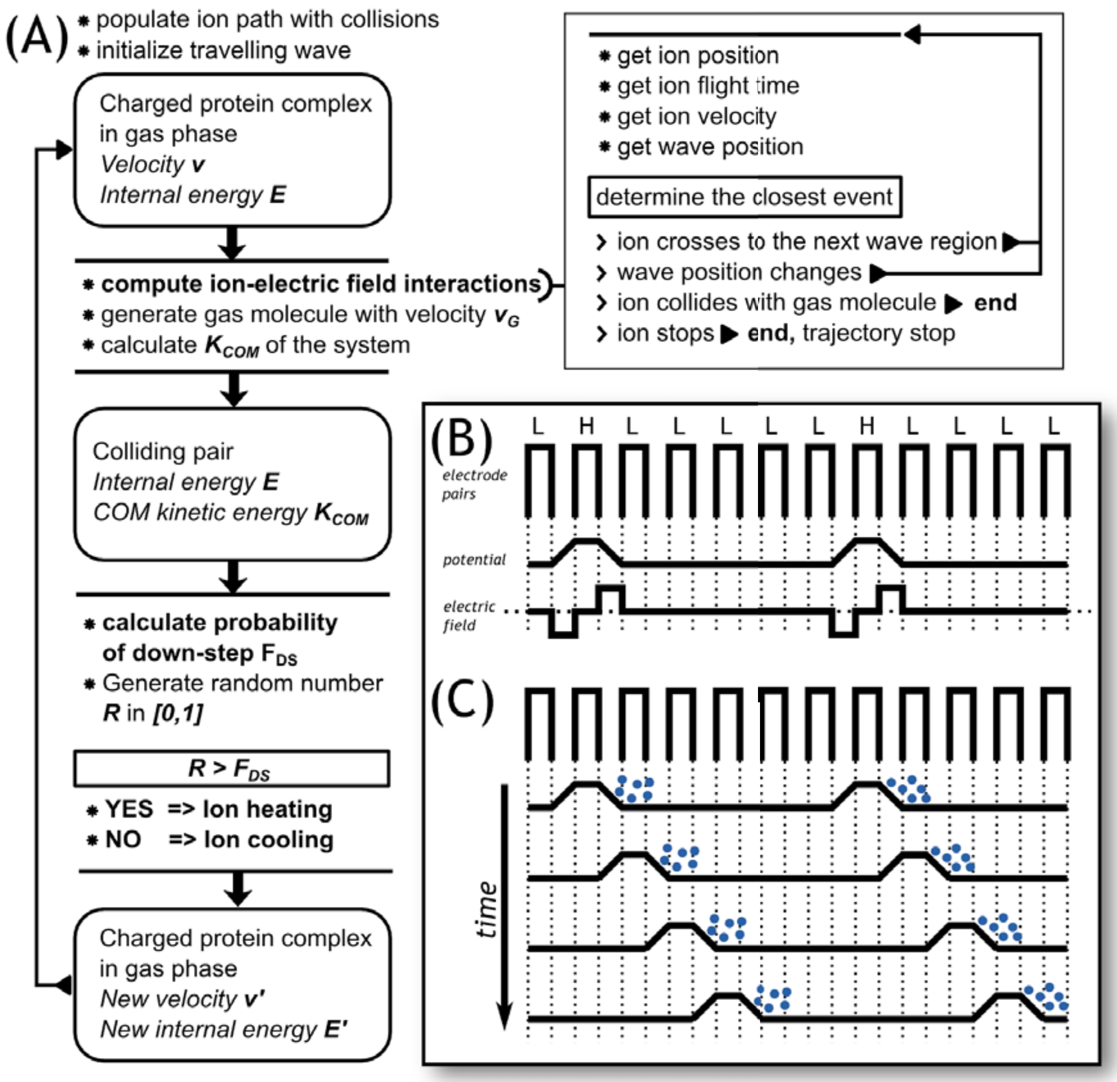


Figure S1 (A) Scheme outlining a single cycle of the Monte Carlo algorithm; (B) diagram showing shapes of electric potential and electric field relative to stacked ring electrodes found in traveling wave ion guides. L is for “low voltage”, H for “high voltage”; (C) schematic representation of electric pulse wave propagating through the traveling wave ion guide and ion pack “surfing” on it.

Following Figure S1 A, the scheme of a unit step of the algorithm can be outlined as follows:

1.1 Ion path is populated with collisions.

Number of collisions is calculated as

$$N_{COL} = \frac{pl\sigma}{k_B T} \qquad \text{(S1)}$$

where *p, l, σ,* k_B and *T* are gas pressure, collision cell length, geometrical collision cross-section of ion with the gas, Boltzmann constant, and temperature, respectively. After, collisions are distributed along the ion path in a quasi-random manner. The ion path is divided into a fixed number of segments, and an equal number of collisions is randomly distributed in each segment, thus ensuring that the overall gas density is maintained along the path.

1.2 Traveling wave initialization.

The traveling wave position is determined randomly from the pool of available positions prior to each run.

1.3 Interactions between the ion and the electric field.

A large variety of linear collision cells are currently used in beam-type instruments, the most typical being RF multipole ion guides [1]. Although the RF field, which serves to confine the ion beam in the plane perpendicular to the main axis, is dynamic and highly complex, the field along the ion path often comprises a simple electric field gradient, which accelerates ions along the cell axis. One exception is a traveling wave technology ion guide (TWIG) [2], where the voltage pulse produced by the array of stacked electrodes is superimposed on the confining RF field. Ion propulsion along the axis is achieved when the ion is “surfing” the wave, i.e. is being pushed in front of it. Here the algorithm to model ion motion along the cell axis in the presence of the traveling wave (TW) is briefly explained. It is described in more details in Section 3. The implementation to a static field should be straightforward.

Algorithm for treating the ion-field interactions in the case of the TW is described schematically in Figure S1. First, ion trajectory parameters, such as position, residence time in the cell and current velocity, are called and its position relative to TW pulses is determined. Four periods of time are then considered: (1) time before the ion reaches the next electrode; (2) time before the wave changes position for the next time; (3) time before the ion collides with the next gas particle; (4) time before the ion stops (when the ion is in the deceleration region of the wave). The next event is chosen on the basis of the shortest of four times. In cases (1) and (2), the trajectory parameters are updated and the check runs again. In case (3), the trajectory parameters are updated and collision is triggered. In case (4), the trajectory is terminated.

1.4 Gas molecule generation.

Gas molecule velocity is generated from the projection of Maxwell-Boltzmann-distributed velocities on the cell axis, resulting in near-normal distribution around zero. Velocity distributions, generated for Ar atoms at 100, 300 and 500 K are shown in Figure. S2.

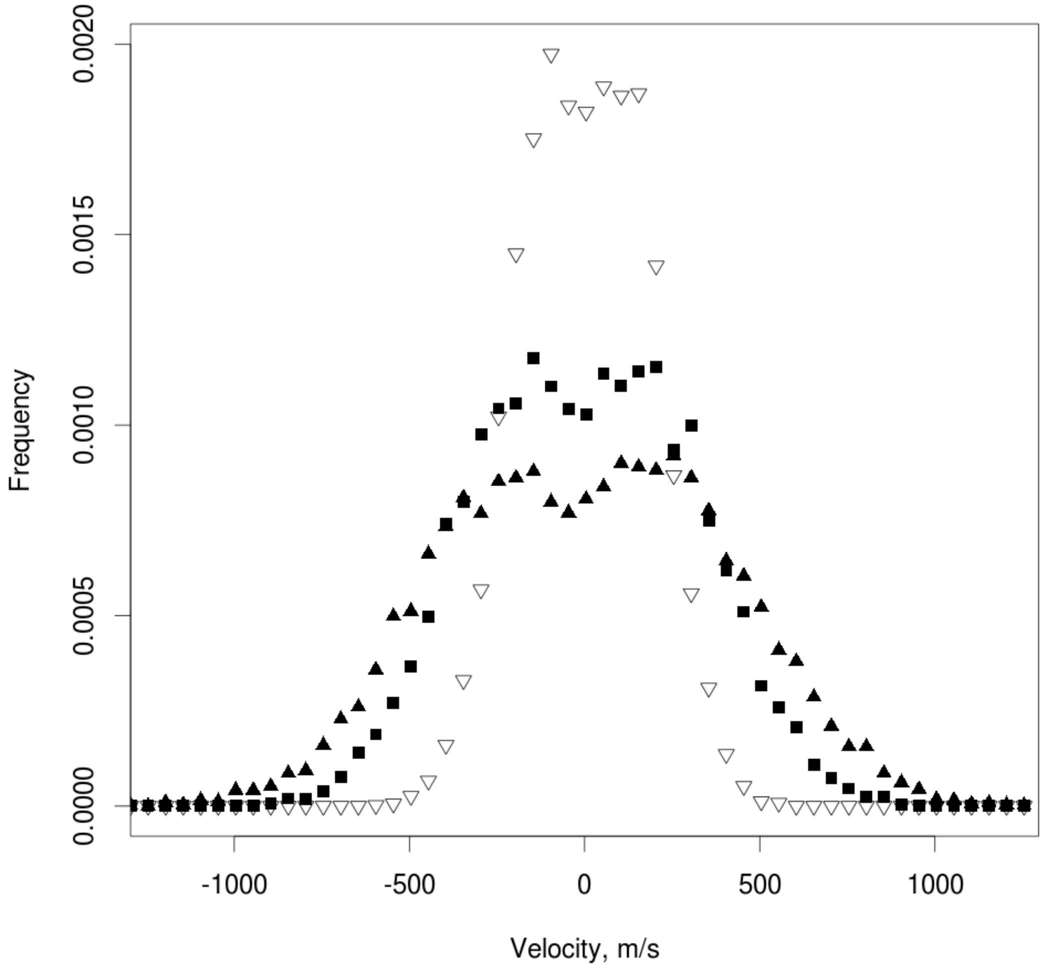


Figure S2. Velocity distribution of gas molecules~~,~~ generated at three temperatures: 100 K (open triangles), 300 K (filled squares) and 500 K (filled triangles)

1.5 Center of mass (COM) kinetic energy calculation.

For a pair of colliding bodies, COM kinetic energy is the maximum energy that can be transformed into internal energy of the system upon collision. For totally inelastic collision, COM kinetic energy after collision is zero, meaning that it was completely transformed into internal energy. COM kinetic energy of a colliding “ion-molecule” pair is the sum of COM kinetic energies of an ion and a gas molecule, and is calculated as follows:

$$K_{COM} = \frac{(\Delta v)^2}{2}\frac{m_I m_G}{m_I + m_G} \tag{S2}$$

where Δv is the difference between velocities of the gas molecule and ion, and m_I and m_G are the masses of the gas molecule and ion respectively.

1.6 Probability of negative energy step calculation.

The probability of a negative energy step is calculated from $P(E,K_{COM},\Delta E)$ - probability density function for the energy step. $P(E,K_{COM},\Delta E)$ is the probability that the system with internal energy E and COM kinetic energy K_{COM} will undergo an energy step change from ΔE to $\Delta E + d(\Delta E)$ upon collision. Details regarding the calculation of the probability density function are given in Section 2.

The probability of a negative energy step is calculated as corresponding cumulative distribution function at $\Delta E = 0$ (eq. S3):

$$W(E,K,0) = \int_{-E}^{0} P(E,K,\Delta E)d(\Delta E) \tag{S3}$$

The cumulative distribution function $W(E,K_{COM},\Delta E^*)$ is equivalent to the probability of the energy step being less than or equal to ΔE^*. Thus $W(E,K_{COM},0)$ is a probability of a negative energy step.

1.7 Random number generation.

A real random number R between 0 and 1 is generated. Here and in other cases when a random number is needed, it is generated using C++ implementation of Mersenne Twister random generator [3]. This is a robust algorithm that provides a highly equi-distributed array of pseudo-random numbers with period of $2^{19937} - 1$.

As proposed in [4], R is then compared with $W(E,K_{COM},0)$. If $R > W(E,K_{COM},0)$ then ion excitation occurs, otherwise, ion cooling is triggered.

1.8 Energy step generation

After the sign of ion internal energy change is known, the energy step is generated. In the case of ion excitation, a positive energy step has to be generated. The ion excitation depends on the COM kinetic energy of the system and is characterized by the energy transfer efficiency η - the average fraction of COM kinetic energy transformed into internal energy of the ion [5]:

$$\Delta E_+ = \eta K_{COM} \tag{S4}$$

η is not a fixed number but rather a function of the COM kinetic energy [6],[7]. The average energy transfer $<\eta>$, calculated for leu-enkephalin, is 12.8 ± 2.1% [8]. To the best of our knowledge, this is the most accurate value reported to date. In the present study, the positive energy step is generated following the exponential function (S5):

$$\Delta E_+ = K_{COM} e^{-\alpha_+ R_+} \tag{S5}$$

where α_+ is an adjustable parameter and R_+ is a real random number between 0 and 1. For the simulations presented in this study, $\alpha_U = 8.05$ was chosen, so that $<\eta>$ would be equal to 12.8%.

A negative energy step is generated in a similar manner (S6). Maximum absolute value of the negative energy step is ΔE_{-0}, for which $W(E,K,\Delta E_{-0}) = R$. Thus R is a probability of the energy step being less or equal to ΔE^*. The exponential distribution function is then applied.

$$\Delta E_- = \Delta E_{-0} e^{-\alpha_- R_-} \tag{S6}$$

where α_- is an adjustable parameter and R_- is a real random number between 0 and 1. α_- should be adjusted so that the system maintains a steady state, i.e. rates of positive and negative energy steps become equal. Values of α_- can differ from system to system and have to be adjusted for each particular case.

1.9 Calculation of new internal energy and ion velocity.

New ion internal energy E' is calculated as $E' = E + \Delta E$.

New ion velocity v_I' is calculated from the laws of energy and momentum conservation as in equation S7.

$$v_I{}' = \frac{m_I v_I + m_G v_G \pm m_G \sqrt{(v_I - v_G)^2 - 2\Delta E * \left(\frac{m_I + m_G}{m_I m_G}\right)}}{m_G + m_I} \tag{S7}$$

where v_I and v_G are velocities of the ion and gas molecule respectively. The sign before the square root is determined by the balance between v_I and v_G: “+” when $v_G > v_I$ and “-” otherwise.

The updated parameters are then passed to point 3 of the algorithm.

2. Collision model

2.1 Theory

The method to determine the cumulative probability function for energy steps over the range of internal and kinetic energies of the ion in gas phase was taken from [4]. We assume that the change in internal energy can occur only as a result of collision, that is, radiative cooling is not considered in this model. This is a reasonable assumption if we consider rates of radiative cooling in the range of 10 s^{-1} , as reported for leu-enkephalin in [8] and that this rate is largely independent of the size of the protein [9]. The time that the ion spends in the collision cell in conventional beam-type mass spectrometers is well below 1 ms, which means that the loss of internal energy as a result of radiative cooling will be less than 1 percent during all the time ion resides in the cell.

The colliding pair —ion and gas molecule in gas phase— is treated as a system that is close to thermal equilibrium. Please note that in Sections 2.1, 2.2 and 2.3 COM kinetic energy is denoted K and not K_{COM} to make the equations easier to read.

Following *Plass et. al.,* we apply a detailed balance principle, stating that in thermal equilibrium rates of positive and negative internal steps are equal. For the system, explicitly characterized by internal energy of the ion and COM kinetic energy of the system ion-molecule, the latter can be written as follows:

$$\rho(E)\, K\sigma(E,K)\, P(E,K,\Delta E) =$$

$$= \rho(E+\Delta E)\, K\sigma(E+\Delta E, K-\Delta E)\, P(E+\Delta E, K-\Delta E, -\Delta E)$$

(S8)

where E is internal energy of the ion, K is COM kinetic energy of the system, ΔE is the energy step, $\rho(E)$ is a density of energy states (DOS) of the ion, $\sigma(E,K)$ is a collision cross-section (CCS) of the ion with respective gas, and $P(E,K,\Delta E)$ is a probability density function for the energy step. The form of $P(E,K,\Delta E)$ for positive energy steps is taken as

$$P(E,K,\Delta E) = C(E,K)\, f(K,\Delta E)$$
(S9)

where $f(K,\Delta E)$ is the probability density for the system with COM kinetic energy K_{COM} to undergo a positive energy step ΔE, and $C(E,K)$ is a normalization coefficient. The further strategy is to construct the equation for $C(E,K)$ that can be solved over all K, E. Combining (1) with the normalization condition, *Plass et. al.* obtain the equation:

$$C(E,K)\int_0^K f\left(f\left(K,\Delta E'\right)\right) d\left(\Delta E'\right) = 1 - \int_{-E}^{0} C\left(E+\Delta E', K-\Delta E'\right)\cdot$$

$$\cdot\frac{K-\Delta E'}{K}\frac{\sigma\left(E+\Delta E', K-\Delta E'\right)}{\sigma(E,K)}\frac{\rho\left(E+\Delta E'\right)}{\rho(E)} f\left(K-\Delta E', -\Delta E'\right) d\left(\Delta E'\right)$$

(S10)

2.2 Numerical solution for integral equation

In order to find a numerical solution for the highly complex equation (S10), the following strategy was followed. Brent's method, proposed by *Plass et. al.,* was discarded because of its slow convergence with not well-behaved functions and difficulty of implementation. A custom method was developed instead.

To reduce complexity, we assign simple names for functions in (3):

(S11)
$$A(K) = \int_0^K f\left(f\left(K,\Delta E'\right)\right) d\left(\Delta E'\right)$$

$$B(E,K,\Delta E) = \frac{K-\Delta E'}{K}\frac{\sigma\left(E+\Delta E', K-\Delta E'\right)}{\sigma(E,K)}\frac{\rho\left(E+\Delta E'\right)}{\rho(E)} f\left(K-\Delta E', -\Delta E'\right.$$

(S12)

Then (S10) becomes

$$C\left(E,K\right)=\left(1-\int_{-E}^{0}C\left(E+\Delta E',K-\Delta E'\right)B\left(E+\Delta E',K-\Delta E',-\Delta E'\right)d\left(\Delta E'\right)\right)\left(A\left(K\right)\right)^{-1}$$

(S13)

C(E,K) is calculated on a grid with grid step *Δ.* Since the change of internal energy of size *ΔE* would always correspond to simultaneous change of COM kinetic energy of size -*ΔE*, the change of internal energy of the system can be represented by moving the point along the diagonal of the *(E,K)* grid. Thus the two-dimensional equation (S13) can be transformed into a set of one-dimensional equations, each representing a diagonal of *(E,K)* grid.

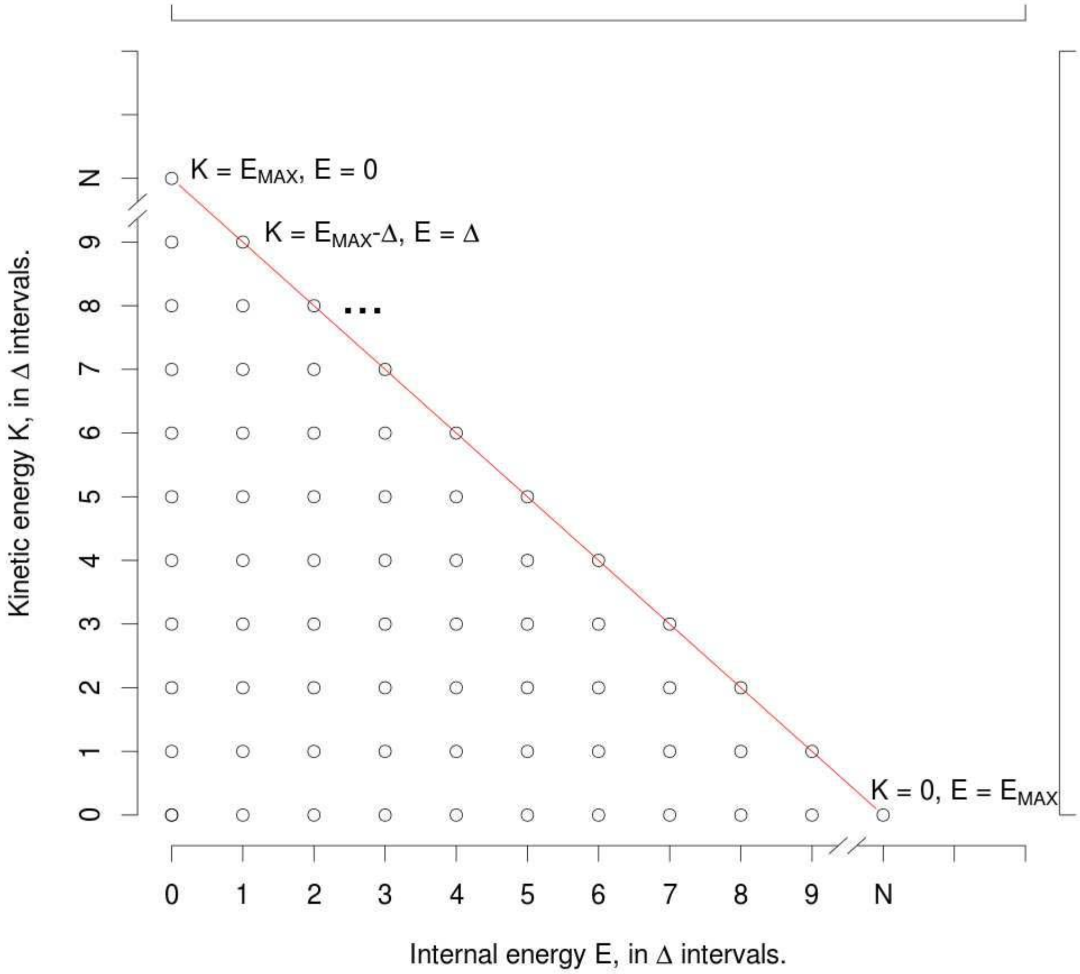


Figure S3. Schematic representation of the numerical solution of integral equation S5.

If we assign a maximum energy that our system can obtain under experimental conditions as E_{MAX}, we can construct a set of one-dimensional equations for each diagonal.

Starting with diagonal from ($K = E_{MAX}$, $E = 0$) to ($K = 0$, $E = E_{MAX}$) (Figure S3), for the left part of equation S13 we can write

$$C\left(E, K\right) = C\left(n\Delta, E_{MAX} - n\Delta]\right)$$

(S14a)

and for unknown part of expression in the right part of equation S13

$$C\left(E + \Delta E, K - \Delta E\right) = C\left(n\Delta + \Delta E, E_{MAX} - n\Delta - \Delta E\right)$$

(S14b)

where *n* is an index of the point within current diagonal. Now, taking into account that any given point *E* and *K* are fixed, *C* now depends only on *ΔE* and index *n:*

$$C_n\left(\Delta E\right) = C\left(n\Delta + \Delta E, E_{MAX} - n\Delta - \Delta E\right)$$

(S14c)

Combining (S14c) and (S13) we obtain

$$C_n\left(0\right) = \left(A\left(E_{MAX} - n\Delta\right)\right)^{-1} \cdot \int_{-n\Delta}^{0} C_n\left(\Delta E\right) B\left(n\Delta, E_{MAX} - n\Delta, \Delta E\right) d\left(\Delta E\right)$$

(S15a)

or

$$C_n\left(0\right) = \left(A\left(E_{MAX} - n\Delta\right)\right)^{-1} \cdot \sum_{i=1}^{n} \int_{-i\Delta}^{-(i-1)\Delta} C_n\left(\Delta E\right) B\left(n\Delta, E_{MAX} - n\Delta, \Delta E\right) d(\Delta E$$

(S15b)

if interval *Δ* is sufficiently small, $C_n(\Delta E)$ in each integration interval *(-iΔ, -(i-1)Δ)* can be approximated using linear interpolation:

$$C_n\left(\Delta E\right) \simeq \frac{C_n\left(-\left(i-1\right)\Delta\right) - C_n\left(-i\Delta\right)}{\Delta} \cdot \Delta E + C_n\left(-\left(i-1\right)\Delta\right) i - C_n\left(-i\Delta\right)\left(i-1\right)$$

(S16)

considering that

$$C_n\left(-\left(n-j\right)\Delta\right) = C\left(j\Delta, E_{MAX} - j\Delta\right) = C_j\left(0\right)$$

(S17)

(S16) can be rewritten as

$$C_n(\Delta E) \simeq \frac{C_{n-i+1}(0) - C_{n-i}(0)}{\Delta} \cdot \Delta E + C_{n-i+1}(0)\, i - C_{n-i}(0)\,(i-1)$$

(S18)

Now we can substitute *$C_n(\Delta E)$* in (S15b) with the expression from (S18):

$$C_n(0) \simeq (A(E_{MAX} - n\Delta))^{-1} \cdot \sum_{i=1}^{n} \int_{-i\Delta}^{-(i-1)\Delta} \left(\frac{C_{n-i+1}(0) - C_{n-i}(0)}{\Delta} \cdot \right.$$
$$\left. \cdot \Delta E + C_{n-i+1}(0)\, i - C_{n-i}(0)\,(i-1)\, B(n\Delta, E_{MAX} - n\Delta, \Delta E)\, d(\Delta E) \right)$$

(S19)

Now we almost have the final expression for the approximate solution of the normalization coefficient *$C(E,K)$*. The final step is to express *$C_n(0)$*, which appears on the right side of the expression in the term of the sum as *$C_{n-i+1}(0)$* when *$i = 1$*. Doing that results in a final equation:

$$C_n(0) \simeq \left[1 + \frac{C_{n-1}}{\Delta} \int_{-\Delta}^{0} - \sum_{i=2}^{n} \left(\frac{C_{n-i+1}(0) - C_{n-i}(0)}{\Delta} \int_{-i\Delta}^{-(i-1)\Delta} \Delta E B d(\Delta E) + \right. \right.$$
$$\left. \left. + (iC_{n-i+1}(0) - (i-1)C_{n-i}(0)) \int_{-i\Delta}^{-(i-1)\Delta} B d(\Delta E) \right] \cdot \left(A + \frac{1}{\Delta} \int_{-\Delta}^{0} \Delta E B d(\Delta E) + \int_{-\Delta}^{0} B d(\Delta E) \right)^{-1}$$

(S20).

Here *$B(n\Delta E, E_{MAX}-n\Delta E, \Delta E)$* is written as *$B$* to reduce the size of the equation.

This equation allows computation *$C_n(0) = C(E,K)$* for any *E, K* along the diagonal, given that the first point, *$C_0(0) = C(0,E_{MAX})$* is known. It can be calculated from (S5) as

$$C(0, E_{MAX}) = \frac{1}{A(E_{MAX})}$$

(S21)

This way *$C(E,K)$* can be computed for any *(E,K)* on the chosen grid.

2.3 Input parameters and probability density function

In order to solve equation (S10), a set of parameters related to the particular system has to be known: DOS of the ion *$\rho(E)$*, CCS of the ion with respective gas *$\sigma(E,K)$,* and probability density for positive energy step *$f(K,\Delta E)$*.

CCS will affect the number and the rate of collisions experienced by the ion. In a general case, CCS depends on the kinetic energy of the ion. However, in case of large biological ions, the effects produced by direct collision between the gas molecule and the ion dominate over the scattering of molecules on the edges of the ion, and geometrical CCS can be used for all kinetic energies. Geometrical CCS can be obtained either computationally, using for example MOBCAL software [10], or from the experiment [11]. For this study, previously published experimentally obtained values were used [12].

Obtaining DOS even for a moderately-sized protein is difficult and computationally expensive [13]. Fortunately, in the equation (S3), DOS is presented as a ratio *ρ(E+ΔE)/ρ(E)*. With analytical approximation for DOS suggested in [14], this ratio for systems close to thermal equilibrium can be safely approximated with a Boltzmann factor (S22).

$$\frac{\rho\left(E+\Delta E\right)}{\rho\left(E\right)} \simeq e^{\frac{\Delta E}{k_B T}} \tag{S22}$$

An appropriate probability density function for the positive energy step *f(K,ΔE)* has to be chosen. Generally, this function has to fulfill the following requirements:

- provide lower probability for higher energy step
- provide higher probabilities for the systems with higher COM kinetic energy

Three function were considered as candidates for *f(K,ΔE)* (eq. S23a,b,c):

$$f\left(K,\Delta E\right) = \frac{K-\Delta E}{K} exp\left(-\frac{\Delta E}{K}\right) \tag{S23a}$$

$$f\left(K,\Delta E\right) = 3.44e^{-\Delta E/0.02K} + 1.4\sqrt{1-\frac{\Delta E}{K}} \tag{S23b}$$

$$f\left(K,\Delta E\right) = \frac{K-\Delta E}{K}\cdot\left(3.44e^{-\Delta E/0.02K} + 1.4\sqrt{1-\frac{\Delta E}{K}}\right) \tag{S23c}$$

In equation S23a, an exponential model is considered for the dependence on energy step *ΔE,* which was shown to work well for small molecules [15]. In S23b,c another, more complex function, is considered, derived from the analytical approximation of energy transfer efficiency, as suggested in [5]. COM kinetic energy at the denominator of exponential function provides higher probability of larger energy steps for the system with higher kinetic energy. In addition, we added a multiplier that effectively lowers the probability density once the size of the energy step approaches COM kinetic energy (S23a,c, Figure S4, A,C). It is worth mentioning that: (1) the function in this form does not contain the transfer efficiency explicitly. For convenience, the transfer efficiency is added in a later step, as it is more flexible and requires fewer computations for optimization; (2) the absolute values of numerical coefficients in equations S23b,c are irrelevant because the function will be normalized during subsequent calculations. However, they are kept as in [5] in order to maintain their value relative to each other.

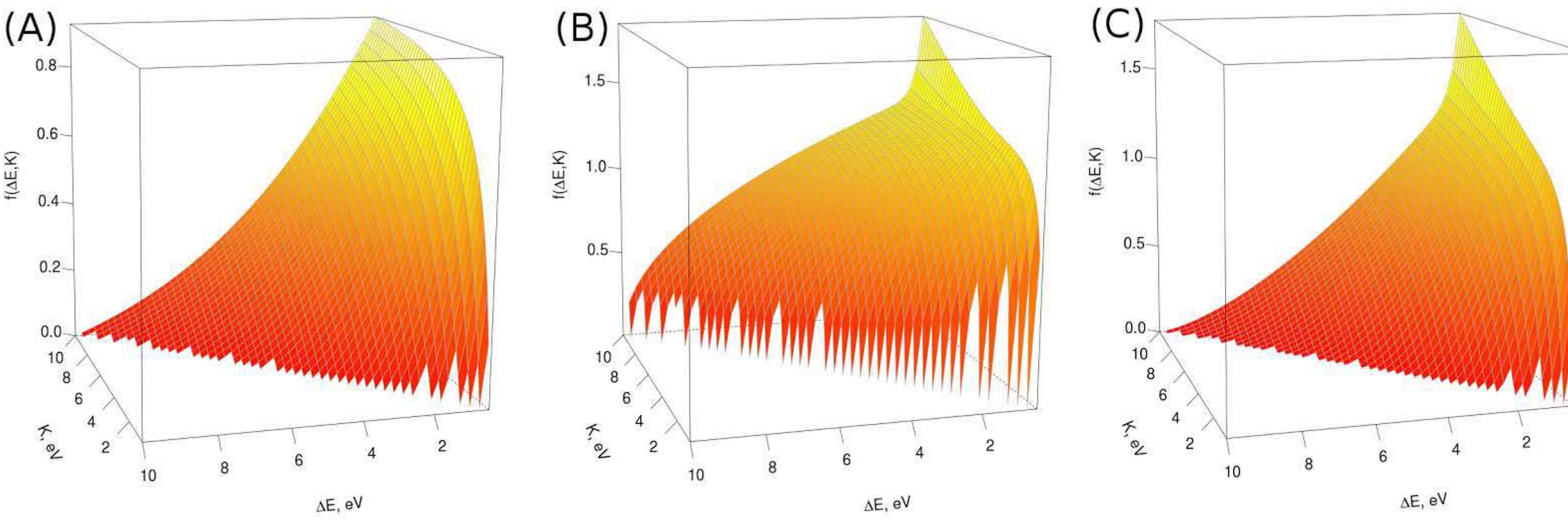


Figure S4. Comparison of surface plots, showing different shapes of probability density for the positive energy step *f(K,ΔE)*. (A) corrected exponential model, eq. S23a; (B) analytical approximation of energy transfer efficiency, eq. S23b; (C) corrected analytical approximation of energy transfer efficiency, eq. S23c;

The performance of these three functions was tested with the experimental data obtained from CID of leu-enkephalin. Function S23b was picked because: (1) the result obtained is closer to the published values [5]; and (2) the probability density function, calculated after using this equation, shows the strongest resemblance with experimental curves, obtained for collisional activation vibrational and rotational degrees of freedom of peptides [6]. However, the difference was minor, and the choice of appropriate model should depend on the system. Equation S23b should work well as a first approximation for peptidic and protein molecules.

After all the input parameters are defined and the equation (S10) is solved for all *(E,K)*, the probability density function can be computed for each *E, K,* and *ΔE* as in equation (S9).

3. Traveling wave ion guide simulation

In the present method, a simplified picture of TWIG is constructed that does not take into account the focusing RF electric field and considers only the wave.

The ion in a TWIG is fully characterized with three parameters: velocity v_I, path from the entrance *d,* and residence time *t*. Information about the wave is stored as an array of logical values, each corresponding to an electrode pair, where TRUE indicates high potential and FALSE low potential. The ion moves on a one-dimensional trajectory along the cell axis in the field-free environment or under a constant electric field. Four events, which can change the ion's environment, are considered, and the period of time until each of them is calculated. The one that will occur first is found and the corresponding algorithm is triggered. The four options are as follows:

1. Ion-molecule collision. The time to the next collision is calculated from the current ion velocity, acceleration and distance to the next collision. After the collision, a new ion velocity is calculated from the laws of energy and momentum conservation.

2. Ion crosses the position of the next electrode. The time to this event is calculated from the current ion velocity, acceleration and distance to the next electrode. After this event, the ion environment may or may not change, depending on the position of the wave.

3. Change of the wave position. The time to this event is calculated by subtracting ion residence time from the time to the next wave position change. After this event, the array that stores wave position has to be updated.

4. Ion stops. Ion loses kinetic energy in ion-molecule collisions and by decelerating in the rear region of the wave pulse (fig. 3 B,C). In the latter case, the time can be calculated from ion current velocity and oppositely directed acceleration. After the ion stops, the trajectory is terminated.

After events 1-3, the position, velocity and residence time of the ion are updated.

4. Experimental data treatment

The typical output of the quantitative CID experiment is an ion collision voltage-dependent breakdown curve [16]. This curve corresponds to the percentage of product ion detected plotted against collision voltage V_{COL}. If we assume at this stage that after passing through the ion transmission guide, the laboratory frame (LF) kinetic energy of ions in the beam is distributed in a normal-like manner around the median value [7], the fraction of the product ion that is not dissociated at voltage V_{COL} can be calculated as

$$\frac{[M]_P}{[M]_{P0}} = 1 - \int_{-\infty}^{V_{COL}} P_n\left(V_{COL}\right) dV_{COL} \tag{S24}$$

where $[M]_{P0}$ is the initial parent ion peak intensity, and $[M]_P$ is a current parent ion peak intensity and $P_n(V_{COL})$ is a normal distribution probability density function, defined by two parameters - median value of V_{COL} $<V_{COL}>$ and standard deviation σ - as follows:

$$P_n\left(V_{COL}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{\left(V_{COL}-<V_{COL}>\right)^2}{2\sigma^2}\right)} \tag{S25}$$

and expression for $[M]_P$ can be found from S24 and S25:

$$[M]_P = [M]_{P0} \cdot \frac{1}{2}\left[1 - erf\left(\frac{V_{COL} - \langle V_{COL}\rangle}{\sqrt{2}\sigma}\right)\right] \tag{S26}$$

where *erf()* is the error function. The experimental breakdown curve can now be fitted to this equation and parameters $<V_{COL}>$ and σ can be found.

These values can serve several purposes. Firstly, simulating ions in collision cells under voltage $<V_{COL}>$ distributed with standard deviation σ can give internal energy of an ion at which half of the complex is decomposed. Secondly, a σ value is needed to estimate the dispersion of kinetic energies. This value is used to generate the distribution of initial ion velocities. Furthermore, $<V_{COL}>$ and σ can be used as input parameters for the optimization run in order to ensure that the system reaches the steady state and remains in it up to the end.

Prior to Monte Carlo simulations of ion trajectories in the collision cell, the cumulative probability density function $W(E, K_{COM}, \Delta E^*)$ has to be calculated for the grid of values E,

K_{COM} between 0 and E_{MAX} (eq. S27). The cumulative distribution function *W(E,K,ΔE*)* is equivalent to the probability of the energy step being less or equal to *ΔE**. Depending on the range of energies and a step of the grid, this part can be relatively computationally expensive. On the other hand, since *W(E,K,ΔE*)* can be calculated independently for each point, it can be easily parallelized.

$$W\left(E, K_{COM}, \Delta E^{*}\right) = \int_{-E}^{\Delta E^{*}} P\left(E, K_{COM}, \Delta E\right) d\Delta E \tag{S27}$$

The values of *W(E,K,ΔE*)* are stored in a series of files that are accessed during Monte Carlo simulations.

Since radiational cooling is not considered in this model, the steady state is maintained when the rate of collisional activation of the ion is equal to the rate of collisional cooling. The average value for the excitation efficiency <*η*> is fixed at 12.8%, so the only parameter that is left for adjustment is *α*. According to [17] , for small molecules the collisional cooling rate is close to the radiative cooling rate and is small in comparison to the excitation rate. Larger molecules undergo more collisions, and collisional cooling rates can be more significant. The *α* value is optimized in a series of short simulations. The trajectory region, where ion velocity is stabilized, is extracted, and the average value of *ΔE* is calculated. When this value is smaller than a certain tolerance, the system is considered to be in a steady state and the corresponding *α* is taken for the production run.

In the next step, a set of simulations is performed at a range of collision voltages that cover the same area as the experimental breakdown curve. For each collision voltage, only the trajectories that are successfully transmitted through the TW collision cell are taken. The average internal energy value of these ions in steady state is calculated and plotted against collision voltage. The curve is then fitted to a power equation (S28) to correlate collision voltages with the internal energy acquired by the ion in the collision cell.

$$E = aV_{COL}^{b} + c \tag{S28}$$

Combining these results with the experimental breakdown curve, the internal energy-dependent breakdown curve is constructed. It corresponds to the percentage of product ions detected after the collision cell plotted against ion internal energy *E*.

In order to extract non-covalent complex dissociation activation energy E_A, the linear region of the curve is found and extrapolated until intersection with the line $[M]_P/[M]_{P0}$ = 1. *E* at the point of intersection is taken as activation energy E_A [16]. This method was chosen because it leaves less uncertainty than simply taking the energy, where a small percentage of complex is dissociated, as activation energy.

The error is estimated as the sum of experimental error and uncertainties of two fittings: *E* vs. V_{COL} fitting and linear fitting of the linear region of the energy-dependent breakdown curve (eq. S29).

$$\delta = \delta_{exp} + \delta_{fit1} + \delta_{fit2} \tag{S29}$$

5. Leucine enkephalin experimental and simulation details

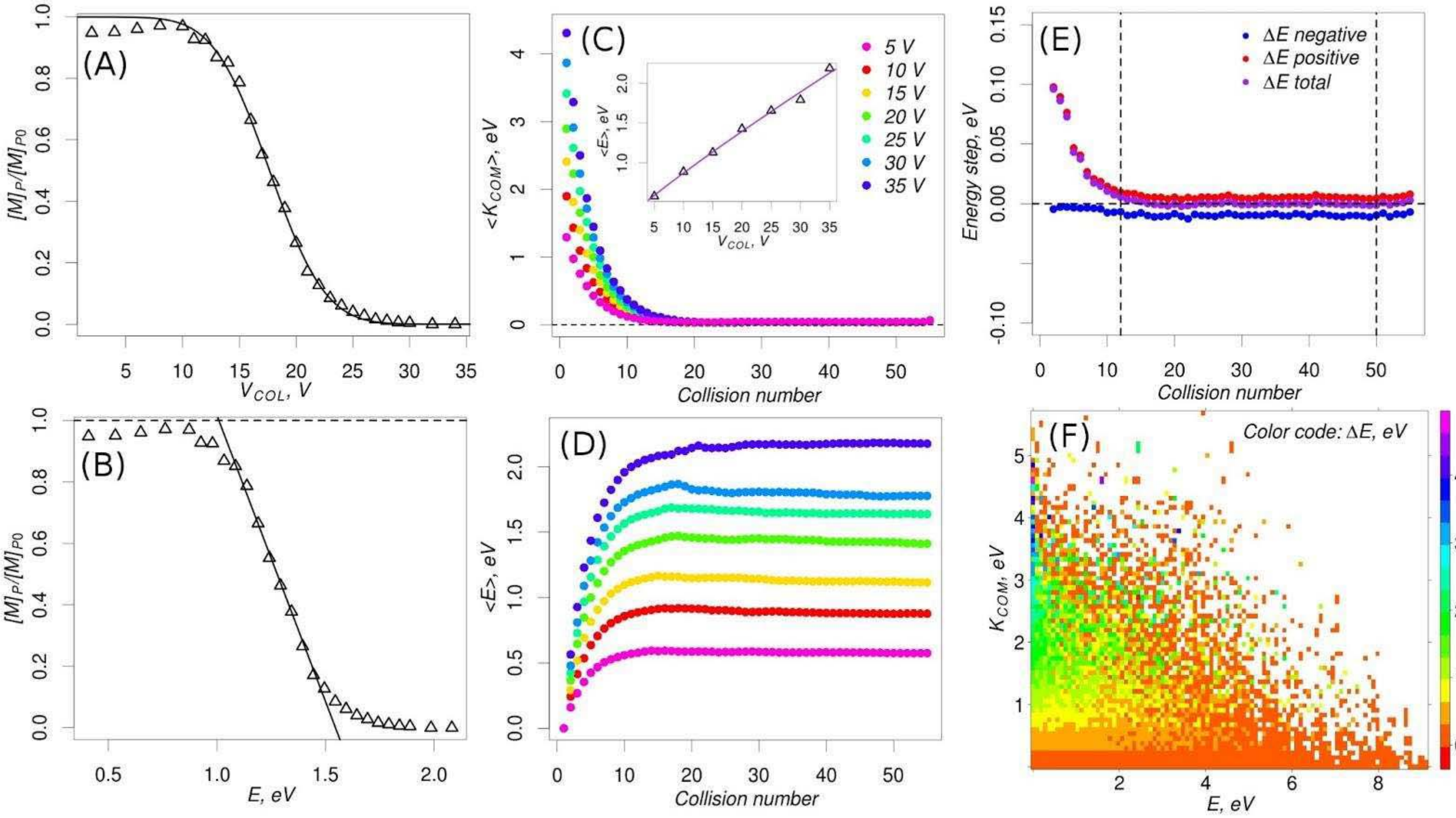


Figure S5. Leu-enkephalin treatment. (A) experimental collision voltage-dependent breakdown curve (empty triangles) and fitted curve; (B) internal energy-dependent breakdown curve and fitting of its linear segment; (C) average COM kinetic energy of ion-molecule system, plotted against number of collisions; (D) average internal energy of the ion plotted against number of collisions; (E) average positive, negative and total energy steps, plotted against number of collisions; (F) level plot of energy steps, generated during seven simulation runs, plotted on a grid of COM kinetic energies and internal energies.

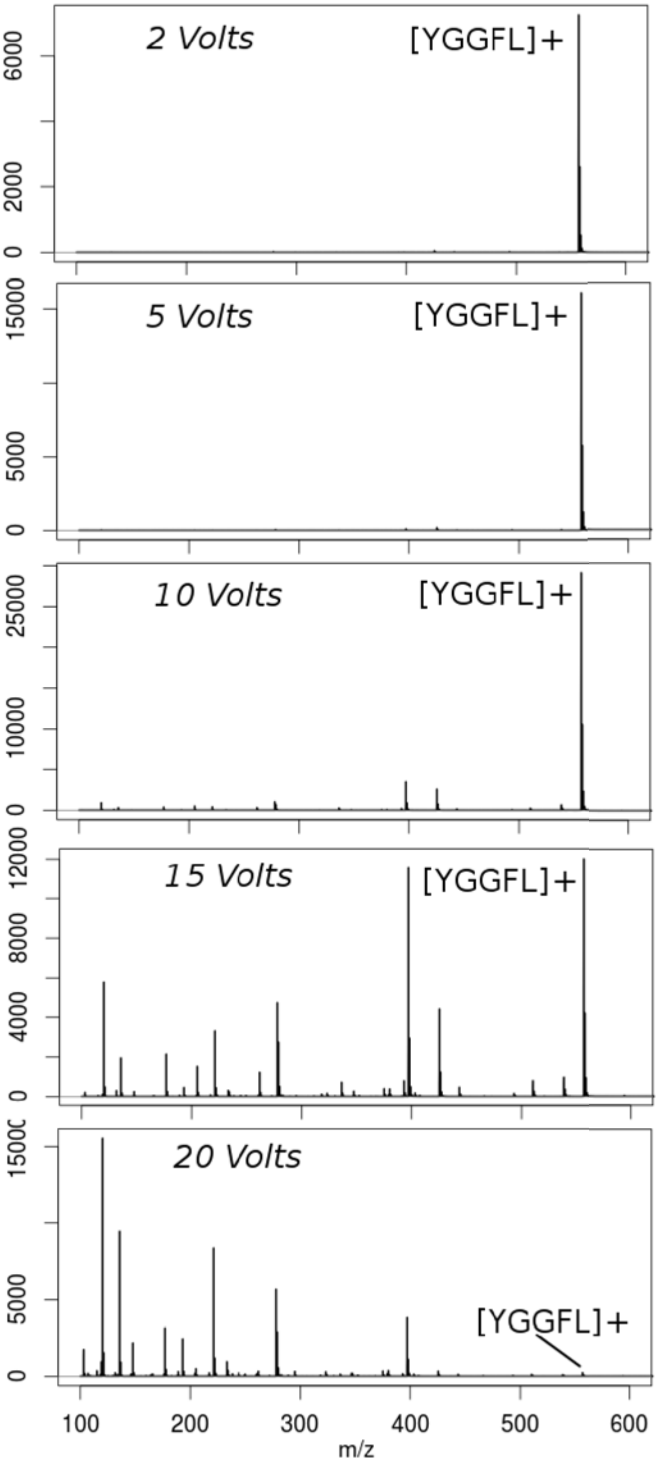


Figure S6. CID spectra of leu-enkephalin at a range of collision voltages.

Experimental V_{COL}-dependent breakdown curves were obtained for 30uM leu-enk dissolved in water/acetonitrile/formic acid 49.5%/49.5%/1% on Waters Synapt G1 instrument, working in the TOF mode (Figure S6). Pressure in the collision cell was $1.4x10^{-2}$ mbar. Spectra were recorded at V_{COL} ranging from 2 to 34 V with increments of 1 or 2 V. (Figure S5 A). Spectra were extracted using MassLynx 4.1 and analyzed using R statistical language [18] with the MALDIquant package [19]. Breakdown curves were calculated following eq. S30:

$$\frac{[M]_P}{[M]_{P0}} = \frac{[M]_P}{[M]_P + \sum [M]_{FR}} \tag{S30}$$

where $\Sigma[M]_{FR}$ is the sum of intensities of all fragments. The fitting of the experimental breakdown curves to eq. S23 is shown in Figure S5 A. $<V_{COL}>$ and σ were found to be 17.66 +/- 0.08 and 3.86 +/- 0.11 respectively.

The next step was to optimize the α_- parameter, as described in section 4. A series of short (400 ions each) simulations was performed with a V_{COL} normally distributed around median value 17.66 with a standard deviation 3.86. A value of $\alpha_- = 12$ was found, at which system maintained a steady state, and the average energy step was 15 ± 1450 μeV. It should be noted that large standard deviation does not invalidate the treatment as long as the overall rates of excitation and cooling are equal. ΔE positive, ΔE negative and ΔE total, averaged across all trajectories and plotted against number of collisions, are shown in Figure S6 E. It can be seen that total ΔE in the region, where ion velocity is stabilized by the TW (in this case, between the 12th and 50th collision), is close to zero, indicating that the system is in a steady state.

After finding α_-, a series of simulations was performed at V_{COL} ranging from 5 to 35 V with 5 V increments. 1000 ions were used in each simulation. The evolution of the average values of the COM kinetic and internal energy of the ions plotted against the number of collisions is shown in Figure S5 C and D, respectively. The behavior of ions in the TW collision cell was modeled at the range of collisional voltages, covering the range of experimental values. The COM collision energy is deposited in the ions' internal degrees of freedom as internal energy. Once the ion velocity is stabilized, its COM kinetic energy is kept constant, maintained by the balance between collisional excitation and de-excitation. Trajectories of ions that successfully reached the end of the cell were extracted and their steady-state internal energy was averaged for each V_{COL}. This value was taken as internal energy acquired by the ion in the TW collision cell. These values were correlated to respective input collision voltages and best fit parameters for the correlation curve (eq. S 24) were found: a = 0.08 ± 0.05, b = 0.89 ± 0.17 and c = 0.26 ± 0.18. The correlation curves and simulated values are shown in the inset in Figure S5 C. A correlation equation was used to construct an internal energy-dependent breakdown curve (Figure S5 B). The linear segment of the curve was extrapolated to meet the intersection with $[M]_P/[M]_{P0} = 1$ line, and the intersection point was found. E at the intersection point, taken as ion dissociation reaction activation energy E_A, was found to be 1.01 ± 0.05 eV. This finding is in good agreement with published values for leu-enkephalin (1.14 ± 0.05 eV) [5]. The discrepancy may arise from the oversimplified treatment of collision cross-section as a constant value, independently of ion kinetic energy. This error is expected to decrease for larger ions.

6. List of VEGF protein ligands

Name	***Sequence***	***$<V_{COL}>$, V***	***σ, V***	***E_A, eV***
P-wt	GGNECDIARMWEWECFERL	24.93 ± 0.11	6.77 ± 0.08	**8.64 ± 0.63**
P-7i	GGNECD**i**ARMWEWECFERL	25.17 ± 0.14	6.84 ± 0.10	**8.31 ± 0.61**
P-18r	GGNECDIARMWEWECFE**r**L	27.64 ± 0.20	6.45 ± 0.14	**9.74 ± 0.71**
P-16f	GGNECDIARMWEWEC**f**ERL	28.00 ± 0.18	6.90 ± 0.13	**9.58 ± 0.70**
P-10m	GGNECDIAR**m**WEWECFERL	34.31 ± 0.10	4.06 ± 0.08	**13.73 ± 1.00**

Table S1. List of peptides used as ligands for VEGF. Lower case bold letters indicate **D** amino acid. For each peptide, median collision voltage, collision voltage standard deviation and calculated activation energy are shown.

References

[1] D. J. Douglas, A. J. Frank, and D. Mao, “Linear ion traps in mass spectrometry.,” *Mass spectrometry reviews*, vol. 24, no. 1, pp. 1–29, 2003.

[2] K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, and R. H. Bateman, “Applications of a travelling wave-based radio-frequency-only stacked ring ion guide.,” *Rapid communications in mass spectrometry : RCM*, vol. 18, no. 20, pp. 2401–14, Jan. 2004.

[3] T. N. Makoto Matsumoto, “Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.”

[4] W. R. Plass and R. G. Cooks, “A model for energy transfer in inelastic molecular collisions applicable at steady state or non-steady state and for an arbitrary distribution of collision energies.,” *Journal of the American Society for Mass Spectrometry*, vol. 14, no. 12, pp. 1348–59, Dec. 2003.

[5] J. Sztáray, A. Memboeuf, L. Drahos, and K. Vékey, “Leucine enkephalin--a mass spectrometry standard.,” *Mass spectrometry reviews*, vol. 30, no. 2, pp. 298–320, 2011.

[6] O. Meroueh and W. L. Hase, “Energy transfer pathways in the collisional activation of peptides,” *International Journal of Mass Spectrometry*, vol. 201, no. 1–3, pp. 233–244, Jul. 2000.

[7] F. Muntean and P. B. Armentrout, “Guided ion beam study of collision-induced dissociation dynamics: integral and differential cross sections,” *The Journal of Chemical Physics*, vol. 115, no. 3, p. 1213, Jul. 2001.

[8] C. Peltz, L. Drahos, and K. Vékey, “SORI excitation: collisional and radiative processes.,” *Journal of the American Society for Mass Spectrometry*, vol. 18, no. 12, pp. 2119–26, Dec. 2007.

[9] R. C. Dunbar, “Infrared radiative cooling of gas-phase ions,” *Mass Spectrometry Reviews*, vol. 11, no. 4, pp. 309–339, Jul. 1992.

[10] A. A. Shvartsburg and M. F. Jarrold, “An exact hard-spheres scattering model for the mobilities of polyatomic ions,” *Chemical Physics Letters*, vol. 261, no. 1–2, pp. 86–91, Oct. 1996.

[11] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, and C. V Robinson, “Ion mobility-mass spectrometry analysis of large protein complexes.,” *Nature protocols*, vol. 3, no. 7, pp. 1139–52, Jan. 2008.

[12] A. Dyachenko, M. Goldflam, M. Vilaseca, and E. Giralt, “Molecular recognition at protein surface in solution and gas phase: Five VEGF peptidic ligands show inverse affinity when studied by NMR and CID-MS.,” *Biopolymers*, vol. 94, no. 6, pp. 689–700, Jan. 2010.

[13] S. Singh, M. Chopra, and J. J. de Pablo, “Density of states-based molecular simulations.,” *Annual review of chemical and biomolecular engineering*, vol. 3, pp. 369–94, Jan. 2012.

[14] K. Hansen, “Comparison of algorithms for the calculation of molecular vibrational level densities.,” *The Journal of chemical physics*, vol. 128, no. 19, p. 194103, May 2008.

[15] A. W. Jasper and J. A. Miller, “Collisional energy transfer in unimolecular reactions: direct classical trajectories for CH4 <--> CH3 + H in helium.,” *The journal of physical chemistry. A*, vol. 113, no. 19, pp. 5612–9, May 2009.

[16] K. Vékey, “Internal Energy Effects in Mass Spectrometry,” *Journal of Mass Spectrometry*, vol. 31, no. 5, pp. 445–463, May 1996.

[17] D. M. Black, A. H. Payne, and G. L. Glish, “Determination of cooling rates in a quadrupole ion trap.,” *Journal of the American Society for Mass Spectrometry*, vol. 17, no. 7, pp. 932–8, Jul. 2006.

[18] R Core Team, “R: A Language and Environment for Statistical Computing.” Vienna, Austria, 2012.

[19] S. Gibb and K. Strimmer, “MALDIquant: a versatile R package for the analysis of mass spectrometry data.,” *Bioinformatics (Oxford, England)*, vol. 28, no. 17, pp. 2270–1, Oct. 2012.

4.

Allosteric mechanisms can be distinguished using structural mass spectrometry

Proc. Natl. Acad. Sci. U.S.A., in press.

Allosteric mechanisms can be distinguished using structural mass spectrometry

Andrey Dyachenko[a,b,c, 1], Ranit Gruber[b,1], Liat Shimon[a], Amnon Horovitz[b,2] and Michal Sharon[a,2]

[a]Departmentsof Biological Chemistry and [b]Structural Biology, Weizmann Instituteof Science, Rehovot Israel [c]Current address: Institute for Research in Biomedicine, Barcelona, Spain [1]These authors contributed equally to this work [2]Correspondence to: Michal Sharon Amnon Horovitz michal.sharon@weizmann.ac.il amnon.horovitz@weizmann.ac.il Tel: 972-8-9343947 Tel: 972-8-9343399

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The activity of many proteins, including metabolic enzymes, molecular machines and ion channels, is often regulated by conformational changes that are induced or stabilized by ligand binding. In cases of multimeric proteins, such allosteric regulation has often been described by the concerted Monod-Wyman-Changeux (MWC) and sequential Koshland-Némethy-Filmer (KNF) classic models of cooperativity. Despite the important functional implications of the mechanism of cooperativity, it has been impossible in many cases to distinguish between these various allosteric models using ensemble measurements of ligand binding in bulk protein solutions. Here, we demonstrate that structural mass spectrometry offers a way to break this impasse by providing the full distribution of ligand-bound states of a protein complex. Given this distribution, it is possible to determine all the binding constants of a ligand to a highly multimeric cooperative system and, thereby, infer its allosteric mechanism. Our approach to the dissection of allosteric mechanisms relies on advances in mass spectrometry, which provide the required resolution of ligand-bound states, and in data analysis. We validated our approach using the well-characterized E. coli chaperone GroEL, a double-heptameric ring containing 14 ATP binding sites, which has become a paradigm for molecular machines. The values of the 14 binding constants of ATP to GroEL were determined and the ATP-loading pathway of the chaperone was characterized. The methodology and analyses presented here are directly applicable to numerous other cooperative systems and are, therefore, expected to usher in a new wave of research on allosteric systems.

chaperonins | cooperativity | Hill coefficient | protein complexes

Multimeric proteins are often subject to allosteric regulation that is achieved by conformational changes induced or stabilized by ligand binding (1). Such allosteric regulation has been described by two classic models: (i) the Monod-Wyman-Changeux (MWC) model (2) in which conformational changes occur in a concerted manner and symmetry is conserved; and (ii) the Koshland-Némethy-Filmer (KNF) model (3) in which conformational changes take place in a sequential manner and symmetry is broken. In addition, it has been proposed more recently that conformational changes can also take place in a probabilistic manner (4). The allosteric control of protein activity is frequently manifested in sigmoidal plots of initial reaction velocity or fractional saturation as a function of the ligand (substrate) concentration that indicate positive cooperativity in ligand binding. It has been impossible, however, to extract any mechanistic insights from these plots (5) since they only show how an average property of the ensemble (e.g. fractional saturation) changes with ligand concentration and do not reveal how the distribution of ligand-bound states changes with ligand concentration. Thus, for example, it is not possible to determine from such sigmoidal plots whether an allosteric transition takes place in a concerted MWC-like fashion (2) or via a sequential KNF-like mechanism (3). This is important since the efficiency of molecular machines is path-dependent and, thus, may depend on the mechanism of allosteric switching. It has also been difficult to determine from such sigmoidal plots how the extent of cooperativity changes with ligand concentration, which may be of importance when there are fluctuations in the physiological concentration of a ligand. Fits of these plots to the Hill equation (6) yield one value of the Hill coefficient, which is a measure of the extent of cooperativity, but the value of the Hill coefficient actually changes with ligand concentration (see *SI Appendix*). The dependence of the Hill coefficient on ligand concentration was shown many years ago in the case of hemoglobin (7), for which it was possible to measure fractional saturation at a range of substrate concentrations that changes by more than four orders of magnitude, but, in general, measurements for such a range of concentrations are not possible. Here, we show that the high resolving power of structural mass spectrometry (MS) allows to determine from a single spectrum the populations of all co-existing states, which differ in the number of bound ligand molecules, thereby facilitating analyses that could not be carried out before of allosteric mechanisms.

Results and Discussion

Our approach is illustrated here for the chaperonin GroEL, which displays intra-ring positive cooperativity and inter-ring negative cooperativity in ATP binding, with respect to ATP (8). Owing to the small difference in mass between apo GroEL (801,312 Da) and its nucleotide-bound states (*SI Appendix* Table S1), MS conditions were first optimized to achieve highly resolved peaks. A series of volatile buffers that are compatible with the structural MS approach were screened for their ability to increase the accuracy of the mass measurement by reducing charge and minimizing the association of water or buffer molecules (Fig. S1). In addition, buffers such as ammonium acetate were avoided since NH_4^+ can substitute for K^+ in promoting ATP hydrolysis by GroEL (9). We discovered that ethylenediammonium acetate (EDAA), a buffer that as far as we know was not used before in MS experiments, gave extremely well resolved peaks, thereby enabling us to distinguish between the different nucleotide-bound forms of the intact GroEL complex (Fig. 1).

MS spectra of 1 μM GroEL, in the presence of 1 mM Mg^{2+}, 200 mM EDAA (pH 7) and different nucleotide concentrations ranging from 0 to 100 μM, were acquired in order to determine the number of bound ATP molecules as a function of ATP concentration. Upon addition of ATP to GroEL, a fine structure is seen within every one of the charge series. Mass measurements of these peaks correspond closely to different numbers of bound

---

Reserved for Publication Footnotes

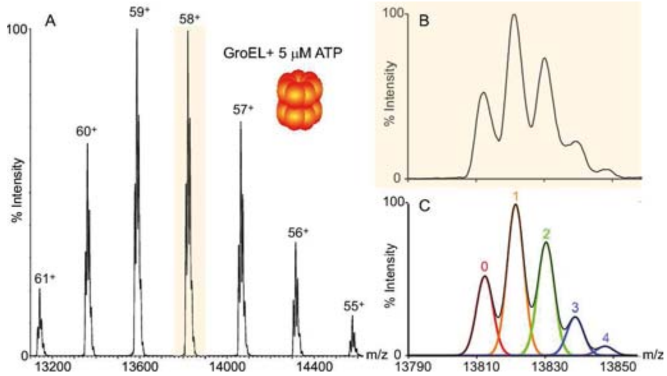


Fig. 1. Determining the distribution in the number of ATP molecules bound to GroEL from the fine structure of each charge state. (A) Mass spectrum of the intact 14-mer GroEL in the presence of 5 μM ATP. The ATP binding number distribution is reflected in the fine splitting of the peaks as highlighted for the 58$^+$ charge state (in orange). (B) A blow-up of the 58$^+$ charge state spectrum. The fine structure of the 58$^+$ charge state is a result of the combination of peaks that correspond in mass to different numbers of bound ATP molecules. The data demonstrate the co-existence of different populations of GroEL/ATP forms. (C) Simulation of a theoretical spectrum assuming co-existence of the following species: (GroEL)$_{14}$, (GroEL)$_{14}$(ATP)$_1$, (GroEL)$_{14}$(ATP)$_2$, (GroEL)$_{14}$(ATP)$_3$ and (GroEL)$_{14}$(ATP)$_4$. A close agreement is seen between the simulated spectrum and the spectrum of the charge state shown in panel B.

ATP molecules with an average mass shift of 534 ± 17 Da, which is in close agreement with the mass of ATP and Mg^{2+} ion (531 Da) (Table S1). This result is shown in Fig. 1 for GroEL in the presence of 5 μM ATP. Each of the charge states of the 14-mer GroEL has an overlay of peaks that are compatible in mass to 0, 1, 2, 3 and 4 bound nucleotide molecules. To assist with the assignment of the peaks arising from this heterogeneous ensemble, we simulated the 58$^+$ charge state of the different ATP-bound GroEL species. It can be seen that the profile of the simulated spectrum fits the observed peaks very closely (Fig. 1*C*). The GroEL spectra acquired for different concentrations of ATP were all assigned in a similar manner. Strikingly, the gradual increase in the number of bound ATP molecules can be clearly observed by superimposing spectra acquired in the presence of increasing ATP concentrations (Fig. 2*A*).

The relative populations of the different ATP-bound GroEL species were determined by measuring the areas of the assigned peaks within each charge state. We took advantage of the multiple charging phenomena displayed in the electrospray ionization method (10) by using the different charge states as replicas (Fig. S2). A drawback, however, of the electrospray method is that the increase in ligand concentration that occurs during the desolvation process of droplets can lead to nonspecific binding (11). We removed the contribution of non-specific binding of ATP to GroEL by applying a mathematical treatment developed previously (12) (*SI Appendix*), which is based on calculating the non-specific binding constant from the MS data, thereby isolating the true ATP binding number distributions of GroEL at different concentrations of ATP (Fig. 2*B*). Given these number distributions, it was possible to calculate the value of the Hill coefficient, n_H, at different degrees of saturation of GroEL's ATP binding sites using Eq. **S6**. The value of n_H is found to first increase with increasing substrate saturation from about 1 until a maximum of 2.33 ± 0.04 and then to decrease back to 1 (Fig. 3*A*). The maximal value reached matches perfectly the single value of n_H of 2.31 ± 0.14 obtained from fitting the plot of fractional saturation determined from the MS data, as a function of ATP concentration, to the Hill equation (Fig. S3). It is also consistent with the value of n_H of 2.68 ± 0.29 determined from fitting to the Hill equation data of initial rates of ATP hydrolysis by GroEL, in the presence of the EDAA buffer, at different ATP concentrations (Fig. S4*A*). The somewhat higher value of n_H determined using the ATPase assay

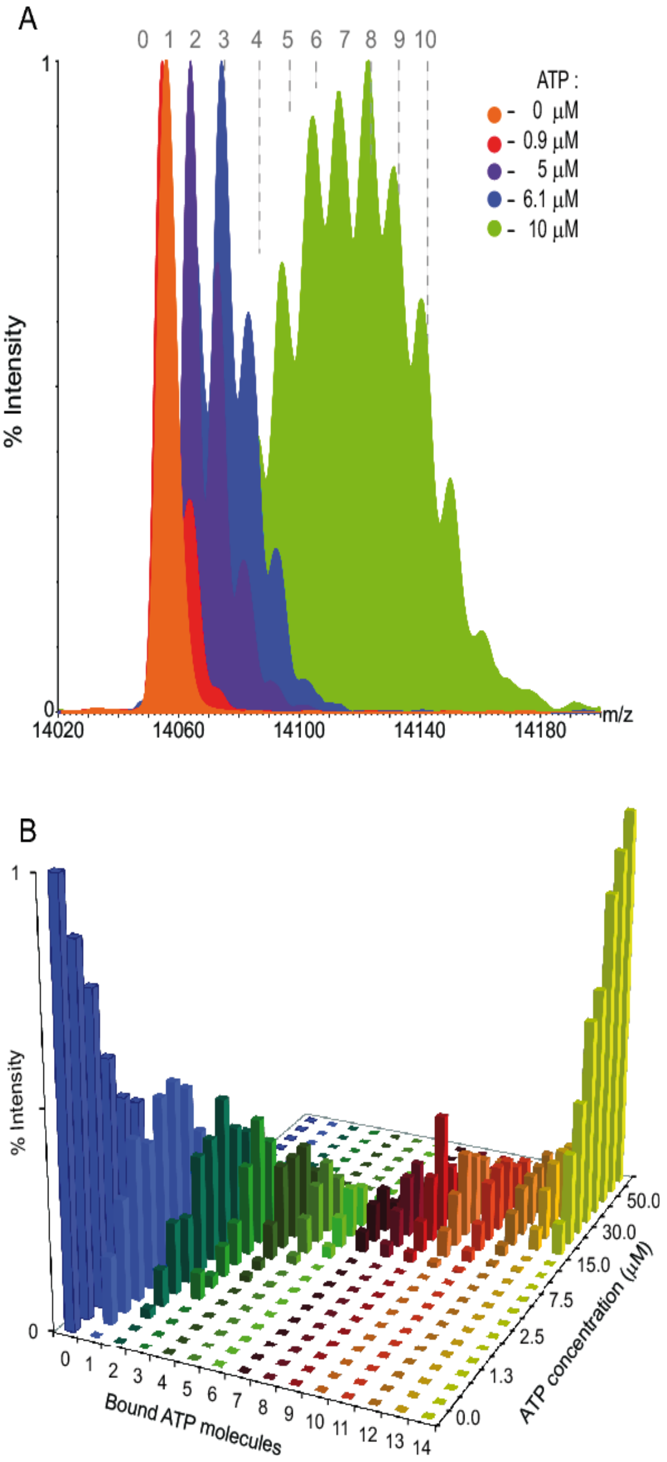


Fig. 2. The relative populations of GroEL molecules with different numbers of bound ATP molecules as a function of the ATP concentration. (A) The spectra of the 57$^+$ charge-state acquired in the presence of different concentrations of ATP are superimposed. The peaks are labeled according to the number of bound ATP molecules to which they correspond. The gradual increase in the number of bound ATP molecules as a function of ATP concentration illustrates the step-wise manner of ATP binding to GroEL. (B)The plot shows the gradual decrease and increase in the populations of apo GroEL and fully bound GroEL, respectively, and the concomitant increase and then decrease in the populations of all the other ATP-bound species.

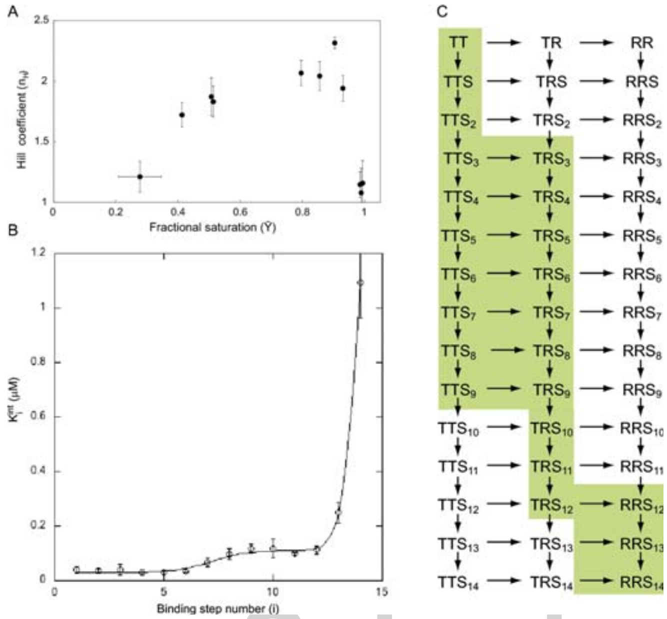


Fig. 3. Dissection of the allosteric properties and pathways of GroEL by using mass-spectrometry. (A) Plot of the value of the Hill coefficient as a function of the fraction of the ATP binding sites of GroEL that are ATP-bound. The values of the Hill coefficient were calculated using Eq. S6 for three charge states separately and then combined to determine the average and standard deviation. The fractional saturation corresponds to the average binding number calculated from the MS data divided by the total number of binding sites. (B) Plot of the change in the value of the intrinsic ATP binding constant as a function of the ligation number. The apparent binding constants (± standard deviations) determined from the MS data were subjected to a statistical correction and then plotted as a function of the binding number. The data were fitted using Eq. S9 in the SI. (C) Scheme showing the ligation pathway of ATP to GroEL. The values of the parameters obtained from fitting the data in panel B were used to calculate the relative populations of the various species in this scheme. The most populated states for each binding stoichiometry are highlighted in green.

is likely to be due to the different K^+ and Mg^{2+} concentrations in the two experiments since these ions are required for hydrolysis but they interfere with the MS measurements. The data in Figs. 3, S3 and S4*A* show that the single value of n_H obtained from conventional Hill plots of degree of saturation (or initial rates) as a function of substrate concentration corresponds to the maximum of the values of n_H determined separately for different degrees of substrate saturation. The increasing and then decreasing value of n_H as a function of degree of saturation is expected from theory (7) as site-site interactions are absent at very low and high substrate concentrations but this has been observed previously only for hemoglobin, to the best of our knowledge, owing to various experimental limitations.

The MS approach described here also allows the determination of the values of the 14 potentially different binding constants of ATP for GroEL. Positive cooperativity in ligand binding implies that the affinity of a ligand increases with increasing ligand occupancy but, in general, the values of the different binding constants for successive binding reactions cannot be resolved using conventional kinetic approaches. Knowledge of the values of the successive binding constants provides a way to validate or rule out potential allosteric mechanisms. In the case of the concerted MWC model (2), for example, the values of all 14 binding constants can be expressed as functions of the equilibrium constant, L, between the low (T) and high (R) affinity states of the protein for the ligand and the respective ligand binding constants, K_T and K_R, of these states. By contrast, no mathematical relationship between the different binding constants is necessarily expected in the case of the KNF model (3). The successive ATP binding constants of GroEL were calculated from the MS data using Eq. S**7** and then corrected using Eq. S**8** to take into account the fact that the number of ways a ligand can bind or dissociate from the protein depends on its ligation state. For example, there are 14 sites to which the first ATP molecule can bind and only one site from which it can dissociate whereas there is only one site to which the 14th ATP molecule can bind and 14 sites from which it can dissociate. Fig. 3*B* shows that the values of the statistically corrected binding constants display a biphasic dependence on the binding number (13). The data in Fig. 3*B* were fitted to Eq. S**9**, which assumes that each ring of GroEL is in equilibrium between T and R states for ATP, in accordance with the MWC model, and that the GroEL double-ring is, therefore, in equilibrium between TT, TR and RR states (8, 13) (Fig. 3*C*). The excellent fit of the data to this equation provides compelling evidence for the concerted nature of the intra-ring allosteric transitions as first assumed in the nested allosteric model for GroEL (8, 13).

The estimates obtained from the fit for the values of K_T, K_R and the allosteric constants for the TT→TR and TR→RR transitions are in line with the lower affinity for ATP of the T state compared with the R state and reflect the strong negative cooperativity between rings. The values of the different parameters are, thus, consistent with the nested model although they differ from earlier estimates (13) because of the different buffer and Mg^{2+} and K^+ concentrations used. A plot of the fractional saturation as a function of ATP concentration that was simulated using the parameter values obtained here yields a monophasic sigmoidal curve, and not a biphasic curve as observed before (13), despite the existence of negative inter-ring cooperativity (Fig S4*B*). Consistent with this simulation, plots of the dependence on ATP concentration of the fractional saturation determined from the MS data (Fig. S3) and of initial rates of ATP hydrolysis, in the presence of the EDAA buffer (Fig. S4), were also found to be monophasic.

The parameters obtained from the fit of the data in Fig. 3*B* were used to calculate the equilibrium relative populations of the species in the scheme shown in Fig. 3*C*. The results show that at least three ATP molecules need to bind to the first ring that switches from T to R in order for its R state to become as (or more) stable than its T state. By contrast, five ATP molecules need to bind to the second ring in order for its R state to become as (or more) stable than its T state, thereby reflecting again the inter-ring negative cooperativity. Overall, our results provide experimental evidence supporting the nested model (13), but they show that switching from the T to R state occurs mostly after binding of 3-5 ATP molecules to the T state.

Conclusions

The MWC model put forward about 50 years ago (2) has been applied to a wide range of systems (14, 15) but the debate between proponents of the concerted (MWC) and sequential (KNF) models has continued. Distinguishing between these allosteric mechanisms is of fundamental interest and is also important because of their functional implications. For example, it has been suggested that eukaryotic chaperonins undergo sequential ATP-promoted allosteric transitions that support domain-by-domain protein folding whereas prokaryotic chaperonins such as GroEL undergo concerted transitions that lead to release and folding in one step (16). The work described here shows that structural MS can be used to determine the relative co-existing populations of a protein with different numbers of bound ligand molecules, thereby allowing one to distinguish between different allosteric models. We anticipate future application of our approach to not only for ATP-driven machines but also for many other protein assemblies that bind cofactors, drugs and other ligands.

Materials and Methods

GroEL expression and purification

Expression and purification of GroEL were carried out as described (17) followed by precipitation in 45% (v/v) acetone (18). Aliquots of purified GroEL were stored at -80 °C in 1 M ammonium acetate buffer (pH 7.5) containing 1 mM DTT.

ATP hydrolysis assays

Initial rates of ATP hydrolysis by GroEL were measured using the phosphate binding protein assay (19) in 200 mM EDAA buffer containing 10 mM $MgCl_2$ and 10 mM KCl at 25 °C. The oligomer concentration of GroEL in these experiments was 250 nM.

Sample preparation for mass spectrometry analysis

Prior to MS analysis, aliquots of 5-6 μM of the intact GroEL complex were thawed and buffer-exchanged three times into the appropriate volatile buffer, suitable for MS analysis, using MicroBioSpin 6 chromatography columns (Bio-Rad). Buffers such as 1 M ammonium acetate, 1 M ammonium acetate supplemented with 10 mM imidazole, 200 mM triethylammonium acetate and 200 mM EDAA were used. 1 mM of Mg^{2+}-acetate was added to all buffer solutions and the pH was adjusted to 7. Following the buffer-exchange step, samples of GroEL (1 μM) were kept on ice and then mixed in a 1:1 ratio with increasing concentrations of ATP ranging from 0 to 200 μM and incubated for 1 min at 25 °C before injection into the mass spectrometer. A MJ Mini thermal cycler (Bio-Rad) was used to ensure accurate temperature control. All reagents were purchased from Sigma-Aldrich.

Mass spectrometry measurements

Nanoflow electrospray ionization (ESI)-MS experiments were conducted using a high mass Q-TOF type instrument adapted for a QSTAR XL platform (20). Conditions were carefully chosen to allow the ionization and detection of GroEL assemblies without disrupting non-covalent interactions. Aliquots of 1.5 μL were electrosprayed from gold-coated borosilicate capillaries prepared in-house as described (21). The following experimental parameters were used: capillary voltage up to 1.2 kV, declustering potential 100-150 V, focusing potential 200-250 V, first and second declustering potentials of 180 V and 0 V respectively. All spectra were calibrated externally by using a solution of cesium iodide (100 mg/ml). Spectra are shown here with minimal smoothing and without background subtraction.

Acknowledgments: .

We thank Drs. Debbie Fass and Uri Alon for helpful comments. M.S. is grateful for the financial support of the Israel Science Foundation (Grant No. 220/10) and a Starting Grant from the European Research Council (ERC) under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement n1 239679. M.S. is the incumbent of the Elaine Blond Career Development Chair. A.H. is an incumbent of the Carl and Dorothy Bennett Professorial Chair in Biochemistry.

1. Perutz MF (1989) Mechanisms of cooperativity and allosteric regulation in proteins. *Q Rev Biophys* 22(2):139-237.
2. Monod J, Wyman J, & Changeux JP (1965) On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12:88-118.
3. Koshland DE, Jr., Nemethy G, & Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5(1):365-385.
4. Martin A, Baker TA, & Sauer RT (2005) Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* 437(7062):1115-1120.
5. Eaton WA, Henry ER, Hofrichter J, & Mozzarelli A (1999) Is cooperative oxygen binding by hemoglobin really understood? *Nat Struct Biol* 6(4):351-358.
6. Hill AV (1910) The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *J Physiol* 40:IV.
7. Wyman J, Jr. (1964) Linked functions and reciprocal effects in hemoglobin: A second look. *Adv Protein Chem* 19:223-286.
8. Horovitz A & Willison KR (2005) Allosteric regulation of chaperonins. *Curr Opin Struct Biol* 15(6):646-651.
9. Viitanen PV*, et al.* (1990) Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are K+ dependent. *Biochemistry* 29(24):5665-5671.
10. Fenn JB, Mann M, Meng CK, Wong SF, & Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246(4926):64-71.
11. Wang W, Kitova EN, & Klassen JS (2003) Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 75(19):4945-4955.
12. Shimon L, Sharon M, & Horovitz A (2010) A method for removing effects of nonspecific binding on the distribution of binding stoichiometries: application to mass spectroscopy data. *Biophys J* 99(5):1645-1649.
13. Yifrach O & Horovitz A (1995) Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. *Biochemistry* 34(16):5303-5308.
14. Changeux JP (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu Rev Biophys* 41:103-133.
15. Changeux JP & Edelstein SJ (2005) Allosteric mechanisms of signal transduction. *Science* 308(5727):1424-1428.
16. Rivenzon-Segal D, Wolf SG, Shimon L, Willison KR, & Horovitz A (2005) Sequential ATP-induced allosteric transitions of the cytoplasmic chaperonin containing TCP-1 revealed by EM analysis. *Nat Struct Mol Biol* 12(3):233-237.
17. Yifrach O & Horovitz A (1998) Transient kinetic analysis of adenosine 5'-triphosphate binding-induced conformational changes in the allosteric chaperonin GroEL. *Biochemistry* 37(20):7083-7088.
18. Voziyan PA & Fisher MT (2000) Chaperonin-assisted folding of glutamine synthetase under nonpermissive conditions: off-pathway aggregation propensity does not determine the co-chaperonin requirement. *Protein Sci* 9(12):2405-2412.
19. Brune M, Hunter JL, Corrie JE, & Webb MR (1994) Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase. *Biochemistry* 33(27):8262-8271.
20. Chernushevich IV & Thomson BA (2004) Collisional cooling of large ions in electrospray mass spectrometry. *Anal Chem* 76(6):1754-1760.
21. Kirshenbaum N, Michaelevski I, & Sharon M (2010) Analyzing large protein complexes by structural mass spectrometry. *J Vis Exp* (40).

Allosteric mechanisms can be distinguished using structural mass spectrometry

Supplementary Information

Andrey Dyachenko[1,2†‡], Ranit Gruber[2†], Liat Shimon[1], Amnon Horovitz[2]* and Michal Sharon[1]*

Affiliations: [1]Departments of Biological Chemistry and [2]Structural Biology, Weizmann Institute of Science, Rehovot Israel

[‡]Current address: Institute for Research in Biomedicine, Barcelona, Spain

†These authors contributed equally to this work

*Correspondence to:

Michal Sharon
michal.sharon@weizmann.ac.il
Tel: 972-8-9343947

Amnon Horovitz
amnon.horovitz@weizmann.ac.il
Tel: 972-8-9343399

Data analysis and correction for non-specific binding

For each mass spectrum, the areas of the peaks corresponding to apo GroEL and its ATP-bound states were calculated using a deconvolution software (peakfit v4, Jandel Scientific, San Rafael, CA). The reliability of the data deconvolution results was validated by comparison of the expected peak position calculated from the mass to charge ratio with the generated peakfit value. After averaging the relative intensities for at least three charge-states, the data were corrected in order to remove the contribution of non-specific binding, as described (1):

$$\frac{C_N}{[E]} = \left(\frac{I_N - I_{N-1}K_n[S]}{I_0}\right)\left(\sum_{j=1}^{\alpha-N}(K_n[S])^{j-1}\right) \qquad [S1]$$

where C_N is the total concentration of the GroEL species with N ATP molecules bound at specific sites, I_N is the intensity corresponding to a population with N ATP molecules bound, [S] is the free ATP concentration, [E] is the concentration of apo GroEL, K_n is the nonspecific binding constant and α designates the largest number of bound ATP molecules that is visible in the spectrum (this number includes the apo state so that if, for example, the highest number of GroEL-bound ATP molecules seen in the spectrum is 5, then $\alpha = 6$). In cases where the peak corresponding to apo GroEL was not present, the following equation was used:

$$\frac{C_N}{[E]} = \left(\frac{I_N}{I_{N-1}} - K_n[S]\right)\left(\sum_{j=1}^{N}[S]^{N-1}K_n{}^{j-1}\prod_{i=1}^{N-j}K_i\right)\left(\sum_{j=1}^{\alpha-N}(K_n[S])^{j-1}\right) \qquad [S2]$$

Analysis of cooperativity

The data corrected for non-specific binding were used to calculate the Hill coefficient as a function of ATP concentration or degree of saturation. The Hill equation is given by:

$$\bar{Y} = \frac{K[S]^{n_H}}{1 + K[S]^{n_H}} \qquad [S3]$$

where $\bar{Y}$ is the fractional saturation, [S] is the ATP concentration, K is the apparent binding constant of ATP to all N sites of the protein and n_H is the Hill coefficient which is assumed to be ATP concentration-independent. This equation was used to fit the data in Fig. S3. Eq. S**3** can be rearranged, as follows:

$$n_H = \frac{\partial \log(\frac{\bar{Y}}{1-\bar{Y}})}{\partial \log[S]} = \frac{[S]}{\bar{Y}(1-\bar{Y})}\frac{\partial \bar{Y}}{\partial [S]} \qquad [S4]$$

where n_H is no longer assumed to be a constant and is a function of [S]. The degree of saturation is given by:

$$\bar{Y} = \frac{\sum_{i=0}^{N} iK_i[S]^i}{N\sum_{i=0}^{N} K_i[S]^i} \qquad [S5]$$

where N is the total number of specific ATP binding sites and i is the number of ATP-bound sites. Combining Eqs. S**4** and S**5** yields:

$$n_H = \frac{\overline{i^2} - (\bar{i})^2}{\bar{i}(1-\frac{\bar{i}}{N})} = \frac{\sigma(i)_{\text{observed}}}{\sigma(i)_{\text{binomial}}} \qquad [S6]$$

In this equation, the Hill coefficient is expressed as the ratio between the observed and binomial standard deviations of the binding numbers. Eq. S**6** was used to calculate the values of the ATP concentration-dependent Hill coefficients from the MS data shown in Fig. 3*A*.

The 14 apparent ATP binding constants of GroEL were calculated from the corrected MS intensities, as follows:

$$K_i = \frac{I_{i(\text{corr})}}{I_{i-1(\text{corr})}[S]} \qquad [S7]$$

where K_i is the apparent binding constant of the ith molecule of ATP (when i-1 molecules are already bound), $I_{i(corr)}$ and $I_{i-1(corr)}$ are the respective MS intensities at a given ATP concentration that correspond to GroEL with i and i-1 bound ATP molecules (after correction for non-specific binding) and [S] is the ATP concentration. Each K_i was calculated at different ATP concentrations and then averaged and the standard deviation determined. The apparent binding constants were converted into intrinsic binding constants by applying a statistical correction that accounts for the number of ways the ith ATP molecule can bind or dissociate from GroEL when i-1 molecules are already bound, as follows:

$$K_i^{app} = \frac{N - i + 1}{i} K_i^{int} \qquad [S8]$$

where N is the total number of binding sites. Given the scheme in Fig. 3*C*, it is possible to express the intrinsic binding constant for the ith site, as follows:

$$K_i^{int} = \frac{K_T^i L_1 L_2 + K_R^i L_2 + K_{R'}^i}{K_T^{i-1} L_1 L_2 + K_R^{i-1} L_2 + K_{R'}^{i-1}} \qquad [S9]$$

where K_T, K_R and $K_{R'}$ are the ATP binding constants of the TT, TR and RR states and L_1 and L_2 are the allosteric equilibrium constants for the TT→TR and TR→RR transitions, respectively (L_1 = [TT]/[TR] and L_2 = [TR]/[RR]). In the derivation of Eq. S**9**, we assumed, for simplicity, that the affinities for ATP of the two different rings in the TR state can be represented by a single binding constant. The values of the parameters obtained from the fit to Eq. S**9** were used to calculate the relative populations in the scheme in Fig. 3*C*.

Reference

1. Shimon L, Sharon M, & Horovitz A (2010) A method for removing effects of nonspecific binding on the distribution of binding stoichiometries: application to mass spectroscopy data. *Biophys J* 99(5):1645-1649.

Table S1. Measured and calculated masses of ATP bound forms of GroEL

Chaperonin complex	Theoretical mass (Da)	Measured mass (Da)[a]
ATP	507	
ATP·Mg^{2+}	531	533 ± 17[b]
GroEL	57,198[c]	57,198 ± 2
$(GroEL)_{14}$	800,766	801,047 ± 10
$(GroEL)_{14}(ATP)_1$	801,298	801,579 ± 16
$(GroEL)_{14}(ATP)_2$	801,829	802,109 ± 29
$(GroEL)_{14}(ATP)_3$	802,361	802,647 ± 47
$(GroEL)_{14}(ATP)_4$	802,892	803,211 ± 14
$(GroEL)_{14}(ATP)_5$	803,424	803,725 ± 21
$(GroEL)_{14}(ATP)_6$	803,955	804,292 ± 31
$(GroEL)_{14}(ATP)_7$	804,487	804,793 ± 26
$(GroEL)_{14}(ATP)_8$	805,018	805,333 ± 40
$(GroEL)_{14}(ATP)_9$	805,550	805,865 ± 36
$(GroEL)_{14}(ATP)_{10}$	806,081	806,403 ± 32
$(GroEL)_{14}(ATP)_{11}$	806,613	806,922 ± 29
$(GroEL)_{14}(ATP)_{12}$	807,144	807,452 ± 34
$(GroEL)_{14}(ATP)_{13}$	807,676	807,986 ± 44
$(GroEL)_{14}(ATP)_{14}$	808,207	808,518 ± 37

[a] Mass errors represent the standard deviation of 4 charge states in 4 different ATP concentrations.

[b] Calculated from the average mass difference between two successive ATP-bound states of GroEL.

[c] The first Met residue is removed.

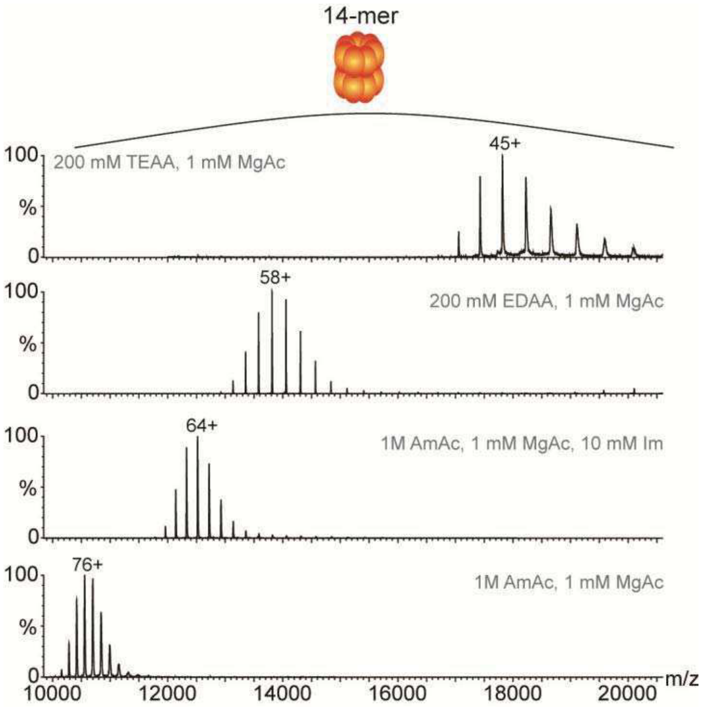


Figure S1. Screening buffer conditions for gaining optimal resolution. Nano-ESI mass spectra of intact GroEL acquired in the presence of either 1M ammonium acetate, 1M ammonium acetate supplemented with 10 mM imidazole, 200 mM EDAA or triethylammonium acetate (TEAA). To ensure cooperative ATP binding 1M MgAc was added to all samples. The presence of 10 mM imidazole lowers the average charge state distribution of GroEL, however, a more pronounced effect that significantly increases the resolution, is observed when EDAA and TEAA were used. While TEAA gave the highest charge state reduction, the peaks were considerably broadened due to its adherence to GroEL's outer surface. Therefore, we used in our analysis the EDAA buffer.

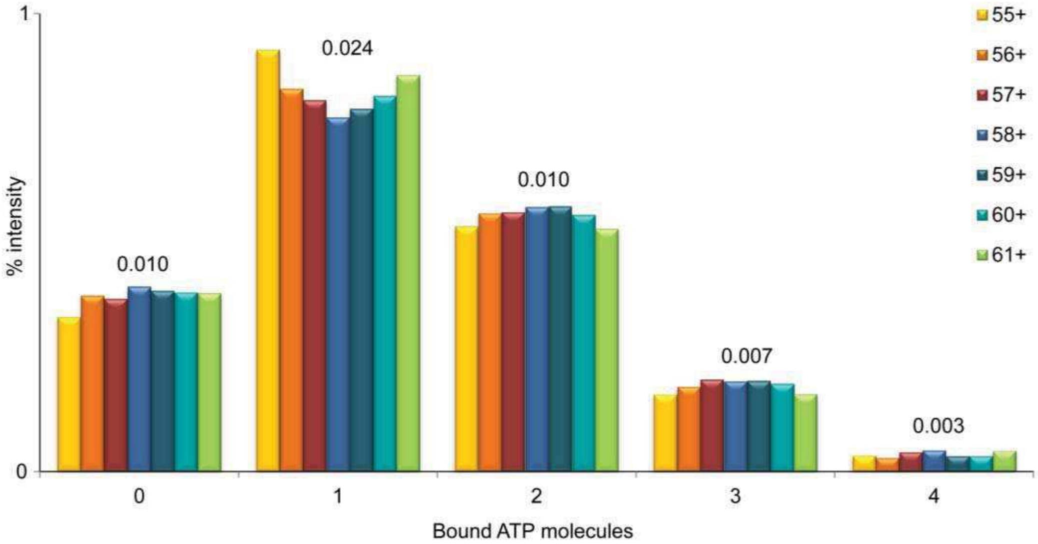


Figure S2. Charge states series give rise to a set of reliable replicas with low standard deviation. The diagram shows the relative populations of GroEL with different numbers of bound ATP molecules determined from the data corresponding to different charge states. The relative populations are remarkably similar among the different charge states with a standard deviation of up to 0.024. The different charge states are color-labeled and the standard deviation values are indicated above the bars corresponding to each population.

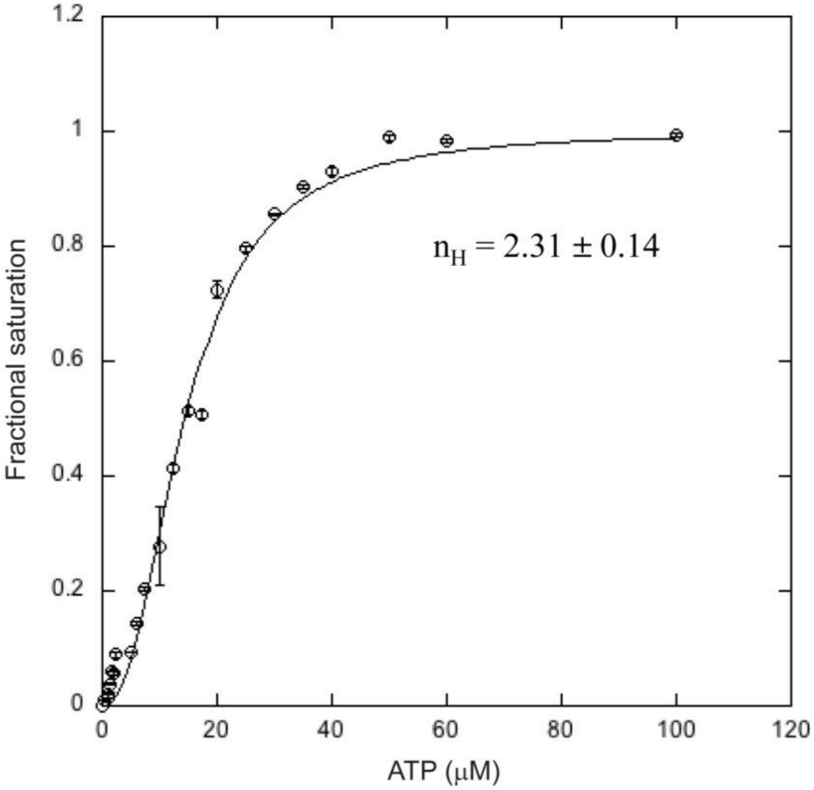


Figure S3. Plot of the fractional saturation determined from the MS data as a function of ATP concentration. The data were fitted to the Hill equation. For further details, see Supplementary Eqs.

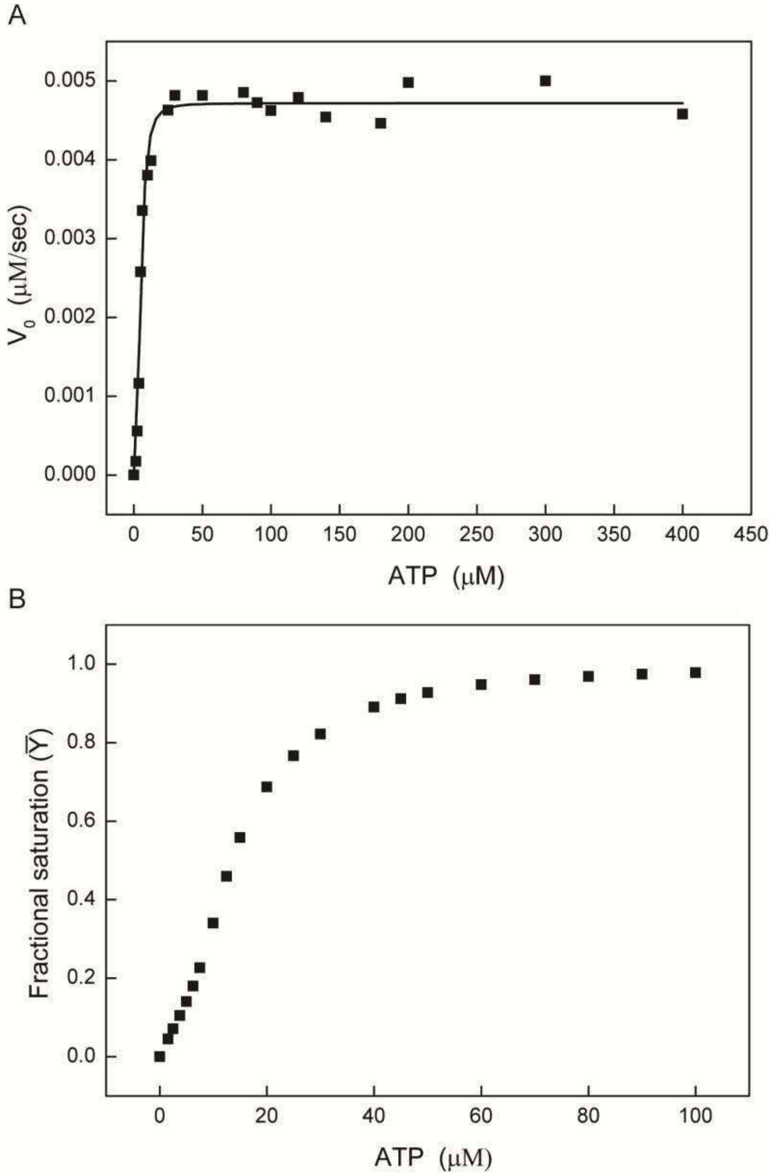


Figure S4. ATP binding and hydrolysis by GroEL, in the presence of the EDAA buffer, displays a mono-phasic dependence on ATP concentration. (A) Initial rates of ATP hydrolysis by GroEL, in the presence of the EDAA buffer, at different ATP concentrations. The data were fitted to the Hill equation. For further details, see Supplementary Equations. (B) Simulation of the fractional saturation of the ATP binding sites in GroEL by different concentrations of ATP using the values of $K_{TT} = 0.03\ \mu M^{-1}$, $K_{TR} = 0.11\ \mu M^{-1}$, $K_{RR} = 1.8\ \mu M^{-1}$, $L_{TT \rightarrow RR} = 3x10^3$ and $L_{TR \rightarrow RR} = 3x10^{15}$ obtained from the fit of the data in Fig. 5 to Eq. S**9**.

5.

Backbone dynamics of vascular endothelial growth factor studied by NMR relaxation combined with normal mode analysis.

Backbone dynamics of vascular endothelial growth factor studied by NMR relaxation combined with normal mode analysis.

(To be submitted to ChemBioChem)

Andrey Dyachenko, Michael Goldflam, Margarida Gairi, Ernest Giralt

ABSTRACT

Proper functioning of biological molecules requires an interplay between the three-dimensional structure of the molecule and its dynamics. Thus characterization of rigid and flexible parts of proteins is crucial for complete understanding of their mechanism of operation. NMR is unique technique in that it is able to provides an atomic resolution study of proteins in the native state. In particular, NMR relaxation can be used to assess conformational flexibility of a protein. In this work we demonstrate that normal mode analysis can efficiently complement NMR relaxation studies and provide information about protein backbone dynamics for the residues that are inaccessible by NMR. We applied this approach to study the backbone dynamics of vascular endothelial growth factor (VEGF) protein, a potent therapeutic target associated with tumor growth. Better understanding of interactions between VEGF and its two receptors can provide means for the development of new drugs, capable of inhibiting the tumor progressing. We used combined data obtained with NMR relaxation and NMR to show that this overly highly rigid protein exhibits high level of flexibility at the receptor binding interface.

---

1. Introduction:

Biological molecules, in particular proteins, are dynamic systems suspended in aqueous and/or lipid media. Their proper functioning requires them to be involved in various processes at different time scales. Different types of motion, characteristic for proteins, include local motions, such as bond vibration (fs - ps), side chain rotation and loop motion (ps - μs), as well as more global processes, such as domain motion, ligand binding, catalysis, folding or allosteric regulation (μs - s) [1],[2] . Although three-dimensional structure of the proteins provided by X-ray or NMR is extremely important for understanding of their function, in many cases it is not enough. It is the interplay between the structure and different types of motion that allows protein to carry out its duties in a correct way. The knowledge of protein dynamics is crucial for our understanding of mechanisms of such events as protein-ligand recognition, allostery, catalysis and protein folding.

2. NMR relaxation

NMR relaxation experiments are used to characterise molecular motion at time scales ranging from picoseconds to seconds [2]. This is done most often by measuring three relaxation parameters: longitudinal relaxation time T_1, transverse relaxation time T_2 and Nuclear Overhauser Effect (NOE). In the protein NMR relaxation experiments ^{15}N and ^{13}C nuclei are usually used. Their relaxation is dominated by CSA and dipolar coupling to the bonded hydrogen atom. Relaxation effects, caused by these interactions can be described mathematically with good precision facilitating analysis of experimental data.

The treatment of experimental data usually involves the following steps: (1) residue-specific values of relaxation times T_1 and T_2 (or relaxation rates $R_1 = 1/T_1$ and $R_2 = 1//T_2$) are determined by fitting the intensities of cross-peaks to single exponential decay equation [3]; (2) the spectral densities $J(\omega)$ are calculated for a specific set of frequencies [4]; (3) conclusions about protein dynamics are drawn.

In case of ^{15}N relaxation an elegant approach can be used to calculate spectral densities, that is called reduced spectral density mapping (RSDM) [5]. Number of simplifications, suggested in [4], allow to construct the following set of equations:

$$J(0.87\omega_H) = \frac{1}{5d^2}\left|\frac{\gamma_N}{\gamma_H}\right|(1 - NOE)R_1 \tag{1}$$

$$J(\omega_N) = \frac{R_1 - 6.244d^2 J(0.87\omega_H)}{3(d^2 + c^2)} \tag{2}$$

$$J(0) = \frac{2R_2 - R_1 - 4.542d^2 J(0.87\omega_H)}{4(d^2 + c^2)} \tag{3}$$

$$d = -\left[\frac{\mu_0}{4\pi}\right]\gamma_H\gamma_N\frac{h}{4\pi r_{NH}^3} \tag{4}$$

$$c = \frac{-\omega_N \Delta\sigma}{3} \tag{5}$$

where ω_H, ω_N, γ_H and γ_N are Larmor frequencies and gyromagnetic ratios of ^1H and ^{15}N respectively, μ_0 is a magnetic constant, h is Planck constant, r_{NH} is the length of NH bond and $\Delta\sigma$ is a CSA of ^{15}N nucleus. With equation 1-5 it is possible to obtain spectral density for “fast” time scale ($J(0.87\omega_H)$ and $J(\omega_N)$, ps-ns time scale) and “slow” time scale ($J(0)$, μs and slower). However, this treatment is associated with fairly strong assumptions about relaxation mechanism. For example, CSA in this approach is considered constant across all 15N nuclei in the protein, whereas it was shown that its distribution around mean value can be quite wide [6],[7] . Another strong assumption that is necessary for RSDM lies in considering the contribution of conformational exchange in transverse relaxation to be negligibly small, which of course is not always true [8]. These assumptions generally reduce the accuracy or RSDM approach.

These problem can be solved (partially or completely) if data obtained at multiple different static magnetic fields is available [9]. The suggested strategy relies on the fact that both c^2 and R_{ex} are proportional to B_0^2. The $J(0)$ is calculated at two field strengths, function of $J(0)$ vs B_0^2 is estimated via least squares fitting and "real" $J(0)$ is calculated at the point where $B_0^2 = 0$.

Another popular approach to the analysis of protein relaxation data is called model-free formalism [10], [11]. This approach takes advantage of the fact that motion of each atom is a superposition of overall slow tumbling of the molecule and its local fast motion. The overall tumbling component is calculated from rotational diffusion tensor and subtracted from the spectral densities. The local motion of each nucleus is then described by two (or more) "model-free" parameters: order parameter S and correlation time of motion τ_e. S (usually treated in the form of S^2) represents degree of restriction of local motion. It scales from 0 to 1, where $S^2 = 0$ corresponds to completely unrestricted motion and $S^2 = 1$ is expected in when there is no free movement at all.

3. Normal mode analysis (NMA)

In order to assess the conformational flexibility of proteins computationally in a rigorous manner, one can consider using molecular dynamics (MD) simulations. However, due to high complexity of the of protein molecules, it often takes extreme amounts of computational time to obtain information about protein behavior on the timescales of tens of microseconds, whereas millisecond timescale dynamics is in most cases unreachable for MD simulations today [12],[13]. There is a number of more simplistic approaches that aim to reduce computational costs of simulations while partly preserving the information, provided by MD. One of the most popular approaches is called normal mode analysis (NMA) [14]. It requires very little computational power, allowing for the assessment of conformational changes of the large biological molecules to be performed within minutes on a desktop computer or even via a webserver [15],[16],[17].

NMA is used to identify the resonant movements of biomolecules (normal modes) and the frequency of this movement (normal frequencies). Depending on the level of delatisation, each atom of the protein or each amino acid residue is represented as a bead. The beads are connected with springs, the elastic properties of which can be derived from simple Hooks's law (elastic network model, ENM) of from the force-field [18]. In the latter case the treatment is called force-field based all-atom NMA (ff-aa NMA) and takes slightly more time, but the moderate size protein can still be treated on a desktop computer in under half an hour [19]. In order to perform ff-aa NMA, one usually needs a dedicated molecular dynamics simulation package, for example, GROMACS [20] or CHARMM [21]. The advantage of aa-ff NMA over the ENM NMA is normal frequencies, estimated with aa-ff NMA are expected to reflect the real frequencies of protein motion. In combination with RMSD, calculated for each atom (or each α-carbon), this can provide a comprehensive picture of time scale and amplitude of protein motion.

4. Vascular endothelial growth factor

Human vascular endothelial growth factor (VEGF) is a signalling protein that is responsible for angiogenesis and vasculogenesis [22]. It plays a central role in generation of blood vessel network during new tissue formation as well as during damaged tissue repair [23]. In addition to that, VEGF is associated with a number of pathological conditions such as tumor growth, cardiovascular diseases, diabetic retinopathy, rheumatoid arthritis and psoriasis [24]. blood vessel growth regulation is performed through binding of VEGF to two receptor tyrosine kinases (RTK), VEGFR-1 (Flt-1), and VEGFR-2 (KDR, Flk-1). Inhibition of these interactions is a validated cancer therapeutic strategy, proved to suppress the development of tumors [25].

Vascular endothelial growth factor A (VEGF-A) is a is a globular covalent homodimer protein, stabilised with two interdomain disulfide bridges. Each monomer is in turn stabilised by three more disulfide bonds organized in a form of rigid cystine-knot motif [26]. In this work a 11-109 construct of VEGF-A121 was used because of its higher solubility in comparison with the full length protein. It was shown to retain full activity with both receptors [27] This construct features mainly β-sheet structure, with one exception being F17-Y25 α-helix at the N-terminus and short helical fragment I35-E38. In addition to that each monomer has two loop regions, D63-G65 (loop 1) and P85-G88 (loop 2). Residues of both loop regions and helix the N-terminal α-helix are involved in interactions between VEGF and its receptors [28], [29],[30].

5. Materials and methods

Protein expression and sample preparation

VEGF A121 11-109 construct was produced in a similar way to that described by Fairbrother et al. [27]. Isotopic labeling of all-^{15}N was performed by using E. coli strain B834 (DE3) (purchased from Cambridge Isotope Laboratories, Inc.). Expression was carried out in M9 minimal media supplemented with 1 g/l $^{15}NH_4Cl$.

Sample preparation

After purification, protein was buffer exchanged into either 50 mM PB, 50 mM NaCl 0.02% NaN_3 in 90:10 H_2O:D_2O, pH 7.0 buffer (Sample 1, 50 mM PB buffer) or 25 mM PB, 50 mM L-Arginine, 50 mM L-Glutamic acid, 0.02 % NaN_3 in 90:10 H_2O:D_2O, pH 7.0 (Sample 2, 50 mM R-E buffer) using PD MidiTrap G-25 columns from GE Healthcare, following standard “gravity” protocol. After that samples were concentrated to 300 μM (Sample 1) and 240 μM (Sample 2) using Centriprep Ultracell YM-10 centrifugal filter units from Millipore.

NMR relaxation measurements

Relaxation measurements included rates of ^{15}N longitudinal (T_1) and transverse (T_2) relaxation and the rate of ^{15}N-^{1}H cross-relaxation measured via steady-state $^{15}N\{^{1}H\}$ nuclear Overhauser effect (NOE). All experiments were carried out at 45 C° on a Bruker Avance III 600 MHz (Sample 2) and Bruker Digital Avance 800 (Sample 1) spectrometers both equipped with a TCI cryoprobe.

15N relaxation measurements were acquired using 2D proton-detected heteronuclear NMR experiments, implementing standard enhanced-sensitivity pulse programs based on Farrow et al [31]. A recycle delay of 1.5 s was employed when recording the T1 and T2 experiments. T1 experiments were performed with nine relaxation delays (s): 0.012, 0.035, 0.070, 0.110, 0.210, 0.310, 0.410, 0.760, 1.010. T2 measurements were taken with seven relaxation delays (s): 0.017, 0.033, 0.050, 0.067, 0.083, 0.100, 0.117. The field strength of the CPMG train was 3.5 kHz and a 0.9 ms delay was used between 180 refocusing pulses. The 1H-15N NOE values were measured from two different experiments (NOE and NO-NOE) recorded in an inter-leaved manner. The NOE spectra was acquired with a proton saturation period of 3 s, preceeded by a 2 s relaxation delay, while the NO-NOE or reference spectra was recorded in absence of proton saturation employing a relaxation delay of 7 s.

Data processing and analysis

Data was processed using NMRpipe suite. Model-free analysis was performed using ROTDIF 7 and DYNAMICS 3 programs. For data analysis and graphical representation R language was used [32].

Normal mode analysis

Normal mode analysis was performed using GROMACS [20] 4.5.2 package, compiled with double precisionm with OPLS-AA all-atom force-field [33]. Prior to the analysis, H++ web server was used to assign protonated states [34]. Energy minimization was performed using steepest descent, conjugate gradient, and ultimately Low-Memory BFGS minimising algorithms. The energy was minimized down to 10^{-8} kJ*mol^{-1}*nm^{-1}. The minimized structure was used for normal mode analysis following standard procedure. Resulting Hessian matrix was diagonalized using GROMACS g_anaeig program and Cα RMSD was extracted using custom scripts.

6. Results and discussion

6.1 NMR Relaxation

In order to check for possible aggregation of the sample, the overall rotational correlation time τ_c was calculated from the T_1/T_2 ratio of residues in the rigid core of the protein [35]. The calculated τ_c values was 10.9 ± 0.3 ns and 10.8 ± 0.9 ns for 600 MHz and 800 MHz set respectively. This is in excellent agreement with the value of τ_c = 11 ns calculated from rotational diffusion coefficient, predicted by HYDRONMR program from VEGF crystal structure [36] . Thus, there is no or very little aggregation of the sample.

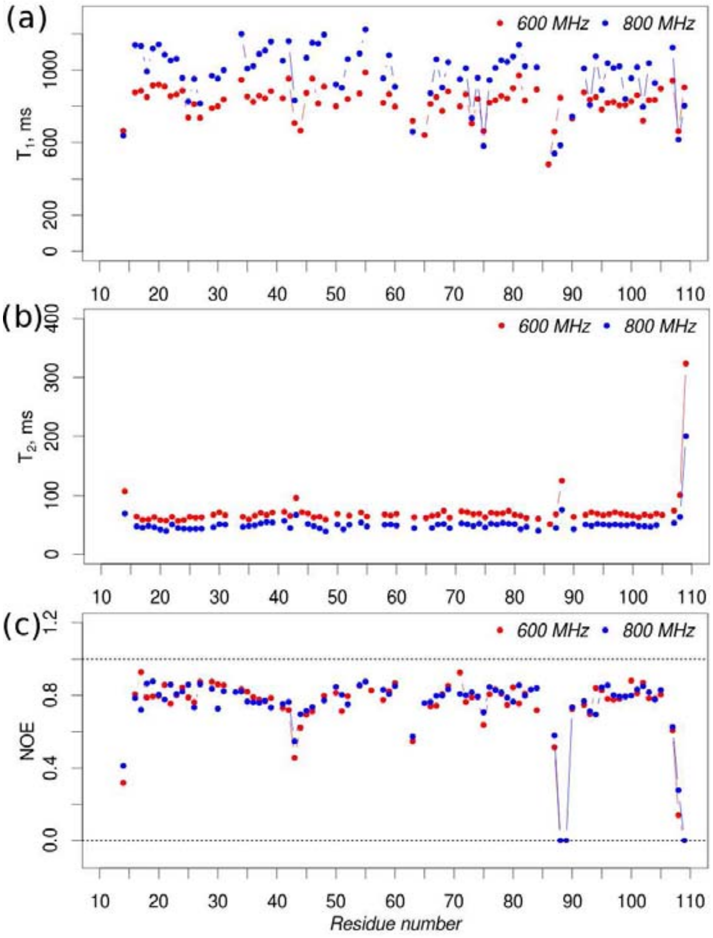


Figure 1. NMR relaxation parameters, acquired at 14.1 T (600 MHz) and 18.8 T (800 MHz); (a) longitudinal relaxation time T_1; (b) transverse relaxation time T_2 (c) ^{15}N-^{1}H NOE.

Two sets of relaxation parameters (T1, T2 and NOE), acquired at 14.1 T (600 MHz proton frequency) and 18.8 T (800 MHz proton frequency) are shown on Figure 1. It was possible to obtain relaxation parameters for 75 to 80% of amino acid residues. Preliminary conclusions can be drawn from the values of T_2 and NOE (Figure 1a,b). Lower T_2 corresponds to the residues involved in conformational changes at μs to ms time scale commonly referred to as chemical exchange. Higher values of T_2 generally indicate elevated backbone dynamics at ps to ns time scale [37]. Low values of NOE also point to regions involved in fast dynamics [38]. In case of VEGF T_2 and NOE show no evidence of chemical exchange, while some residues exhibit higher degree of ps-ns scale dynamics. Namely, D63 and Q87-G89, situated on loop 1 and loop 2 respectively, show low NOE and/or high T_2. In particular, the values of NOE for G88 and Q89 are zero, indicating a very high flexibility of loop 2 on the level of highly disordered terminal residues. In addition to that, NOE and T_2 also indicate ps-ns motion of I43 residue, located in an unstructured region of VEGF.

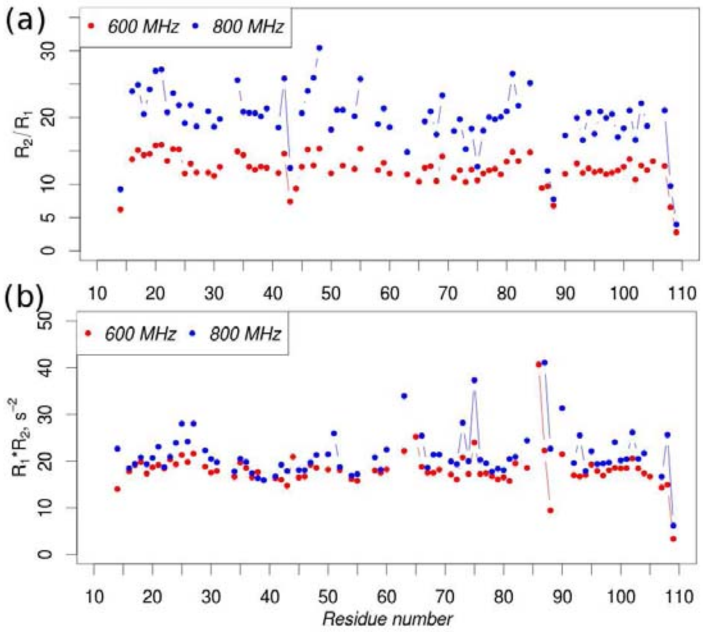


Figure 2. Identification of residues, involved in chemical exchange; (a) higher R_2/R_1 ratio could show if there is a substantial contribution of R_{ex} in R_2; (b) higher values of R_1*R_2 product can provide information similar to R_2/R_1 and at the same time suppress the contribution of motional anisotropy.

In order to identify residues involved in chemical exchange processes different strategies are suggested. Elevated values of R_2/R_1 (or T_1/T_2) ratio is an indication of the presence of chemical exchange because chemical exchange rate constant R_{ex} only contributes to measured R_2 values [38],[39]. Similarly, increase of R_1*R_2 product is also associated with chemical exchange. Moreover, it was shown that R_1*R_2 suppresses the contribution of motional anisotropy, that affect R_2/R_1 ratio [40]. Residue-specific values of both R_2/R_1 *and* R_1*R_2 are shown on Figure 2. Neither plot shows definitive evidence for presence of chemical exchange processes in any region of VEGF. The only residues that could have a contribution of Rex according to R_1*R_2 are H86 and Q87. However, is not confirmed by the two fields in neither of the cases.

In order to obtain more complete picture of backbone dynamics of VEGF, RSDM analysis was performed. Results are shown on Figure 3a,b,c. $J(0.87\omega_H)$ represents intensity of fast motion of backbone amide at frequencies around 700 MHz and 520 MHz for 18.8 T and 14.1 T static magnetic field respectively. $J(0.87\omega_H)$ does not require any prior knowledge about R_{ex} or CSA and can be computed for each residue directly from NOE and T_1 (eq. 1). Not unexpectedly, $J(0.87\omega_H)$ confirms the predictions obtained from NOE values and shows high level of flexibility at the same regions. loop 1 (D63) and 2 (Q87-G89), as well as unstructured region around residues I43 and E44 consistently show higher flexibility. In addition to them, residue N75 showed $J(0.87\omega_H)$ slightly higher than average.

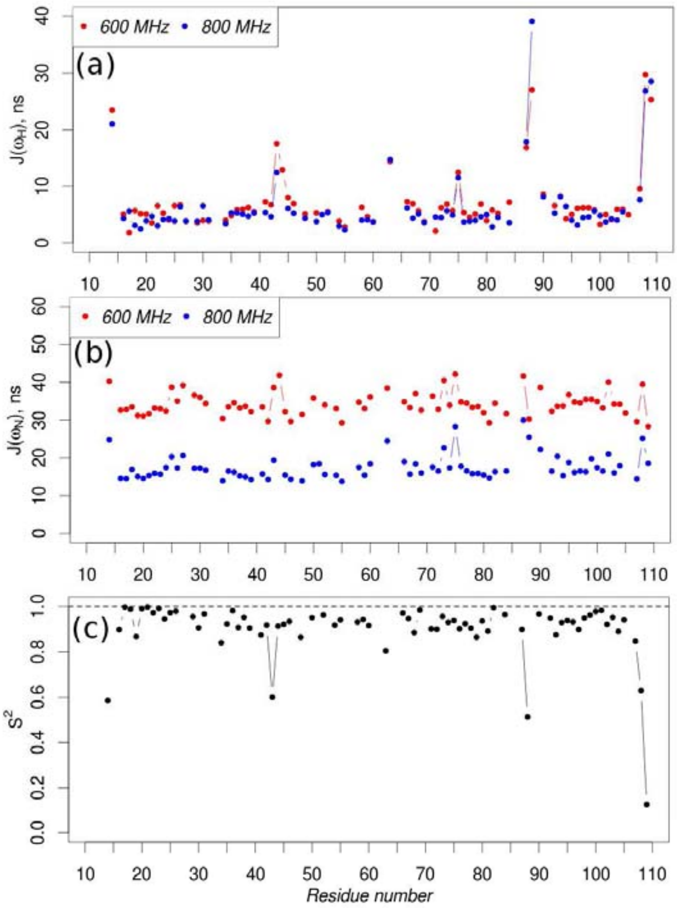


Figure 3. Results of spectral density mapping and model-free analysis of VEGF; (a) and (b) Spectral density mapping of VEGF backbone at ω_H and ω_N respectively; (c) generalized order parameter, derived from model-free analysis of relaxation data acquired at 14.1 T.

The treatment is more complicated with $J(\omega_N)$ and $J(0)$ since in order to accurately calculate them, one has to know site-specific values of CSA (for $J(\omega_N)$) or both CSA and R_{ex} (for $J(0)$). On the other hand, the data analyzed above points on the absence of the chemical exchange processes. This allows to assume that contribution of chemical exchange to the transverse relaxation on observer ^{15}N nuclei is negligibly small and $R_{ex} \approx 0$. Then, if the data acquired at different static magnetic fields is available, residue-specific values of CSA and $J(0)$ can be found. Equation 3 can be rewritten as

$$2R_2 - R_1 - 4.542d^2J(0.87\omega_H) = \omega^2\frac{4}{9}J(0)(\Delta\sigma)^2 + 4d^2J(0) \quad (6)$$

In equation 6 CSA ($\Delta\sigma$) and *J(0)* are independent of static magnetic field B_0, while $\omega_N = \gamma_N B_0$. So if relaxation rates from at least two different B_0 are available, linear fit of eq.6 can provide corresponding values of CSA and *J(0)*. The CSA can then be used to calculate $J(\omega_N)$.

This approach is very sensitive to difference in the experimental conditions between two acquisitions. Unfortunately, the difference in the buffer that exists between two data sets used for this work seems to be enough to render this technique ineffective: it fails to predict reliable values of CSA. So here we used conventional approach to calculate $J(\omega_N)$, i.e, assuming uniform CSA = -160 ppm. The results are shown on Figure 3b. $J(\omega_N)$ reflects motion at frequencies around 80 MHz and 60 MHz for for 18.8 T and 14.1 T fields respectively. Residues E44 (loop 1), D63 (unstructured region), Q87, G88 and H90 (loop 2), and again E73 and N75 (β-sheet 5) exhibit more flexibility than the bulk of the protein.

Another way to assess overall backbone flexibility of the protein molecule is model-free analysis. The plot of squared generalized order parameter S^2, calculated from 600 MHz data set using Figure 3c. The fitting of model-free parameters was performed using DYNAMICS program [41] with axially-symmetric rotational diffusion tensor. ROTDIF 7 [42] program was used to fit the rotational diffusion tensor.

The average value of S^2 calculated for 20 structured globular proteins is 0.839 ± 0.106 [43]. VEGF appears to be a rigid protein with average S^2 = 0.90 ± 0.13. The residues that show lower degree of restriction are again I43, D63 and G88, pointing at three flexible zones, identified before.

6.2 Normal mode analysis

Even with the best theoretically possible 15N-1H NMR relaxation experiment it is impossible to obtain information about proline residues because of lacking backbone amide proton. In real experiments the information is often lost because of signal overlapping or low signal intensity caused by intermediate chemical exchange effects [44]. NMA can serve as complementary technique to NMR relaxation providing insight into mobility of missing residues. RMSD values calculated for α-carbons represent the degree of backbone displacement relatively to the mean value due to the normal mode motion in a residue-specific way. It can be directly compared with the NMR relaxation data.

Level plot (Figure 4a) represents Cα-RMSD calculated for first 40 normal modes of modes of VEGF. These values, calculated for the first 40 normal modes for α-carbons of VEGF backbone are shown on Figure 4. The levelplot shows that different modes correspond to the flexibility of different parts of VEGF backbone, but general trend reveals four zones of higher flexibility. The observation is confirmed by the plot of RMSD (Figure 4b). Loops one and two are not unexpectedly among the most flexible parts. In addition to that, unstructured region around P40 and E72 (β-sheet 5) show higher flexibility.

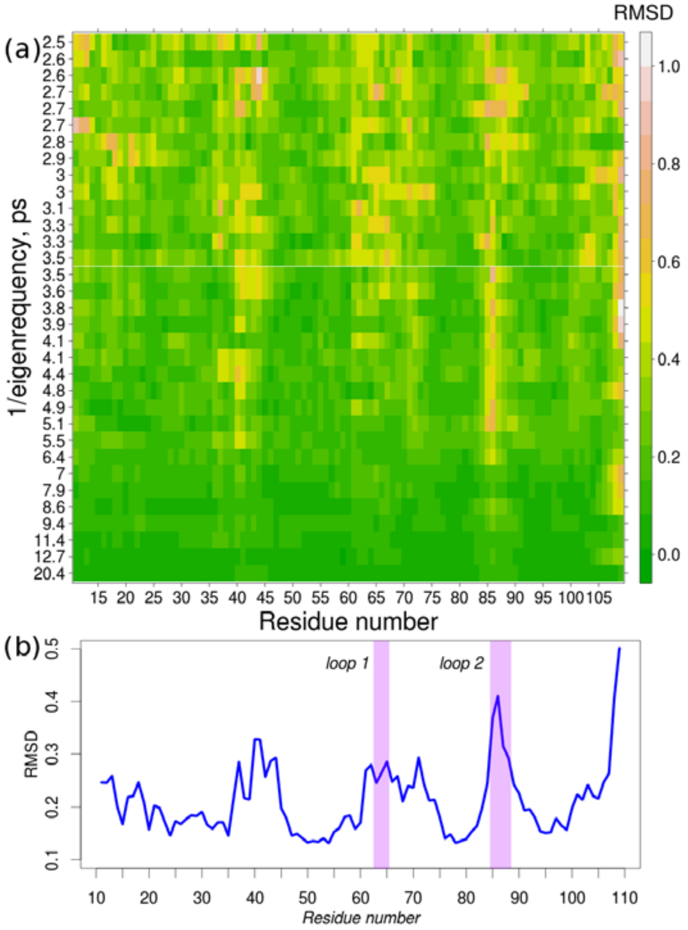


Figure 4. Normal mode analysis of VEGF; (a) level plot of normalized Cα RMSD for modes 7-40 as a function of inverted eigenfrequency; (b) normalized Cα RMSD for averaged over modes 7-40.

6.3 Discussion

Comparison of results, obtained by NMR relaxation and NMA (Figure 5, blue and red bars respectively) show that both techniques identified four flexible regions of VEGF backbone, excluding two termini. D63-G65 (loop 1) and P85-G88 (loop 2) as well as surrounding low-structured regions were identified by both techniques. In addition to that, part of β-sheet 5 and preceding low-structured region T71-N75 and the region around I43 and E44 were also picked up by both techniques. The difference between them often (but not always) arise from incomplete sequence coverage (Figure 5, black bar) provided by NMR. This is the case with both loops, where information of many residues is missing probably due to the conformational motions in the intermediate exchange regime. One exception is N75 residue, that was identified by NMR Relaxation but not by NMA. This is probably due to the fact that first normal modes take into account large overall motion of the protein. Thus it can point out to the flexible region of the protein, but doesn't reach single residue resolution.

In addition to that, comparison of the patches, involved in in interactions between VEGF its two receptors (KDR and Flt-1) (Figure 5, purple bar) show that they often overlap with flexible regions completely or partially (loops 1 and 2 respectively). In case of residues Y46 and K48, they are very close to third identified flexible region. This suggests that flexible patch on the surface of otherwise very rigid protein plays important role in its interactions with receptors and regulation of angiogenesis.

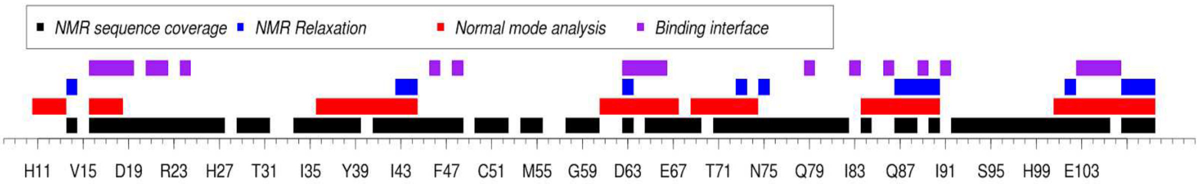


Figure 5. Superposition of flexible regions of VEGF backbone, identified with NMR relaxation and NMA, VEGF-receptor interactions interface. Black bar shows residues for which full set of relaxation parameters was available at least at one magnetic field.

7. Conclusions

In this work we studied backbone dynamics of VEGF by NMR relaxation and all-atom force-field based normal mode analysis. We showed that NMA, being fast and computationally inexpensive technique, can still provide information that confirms and complements NMR relaxation studies.

Four flexible patches were identified in the surface of VEGF, including two loops, unstructured region and part of β-sheet 5. Comparisons of flexible regions with binding interface between VEGF and two receptors (KDR and Flt-1) revealed that three of them overlap with binding interface and one is situated in the immediate vicinity of it. This suggests that increase of flexibility plays role in VEGF-receptor interactions.

References:

[1] A. G. Palmer, "NMR characterization of the dynamics of biomacromolecules.," *Chemical reviews*, vol. 104, no. 8, pp. 3623–40, Aug. 2004.

[2] K. Henzler-Wildman and D. Kern, “Dynamic personalities of proteins.,” *Nature*, vol. 450, no. 7172, pp. 964–72, Dec. 2007.

[3] J. W. Peng, V. Thanabal, and G. Wagner, “2D heteronuclear NMR measurements of spin-lattice relaxation times in the rotating frame of X nuclei in heteronuclear HX spin systems,” *Journal of Magnetic Resonance (1969)*, vol. 94, no. 1, pp. 82–100, Aug. 1991.

[4] N. Farrow, O. Zhang, A. Szabo, D. Torchia, and L. Kay, “Spectral density function mapping using 15N relaxation data exclusively,” *Journal of Biomolecular NMR*, vol. 6, no. 2, Sep. 1995.

[5] R. Ishima and K. Nagayama, “Protein Backbone Dynamics Revealed by Quasi Spectral Density Function Analysis of Amide N-15 Nuclei,” *Biochemistry*, vol. 34, no. 10, pp. 3162–3171, Mar. 1995.

[6] D. Fushman, N. Tjandra, and D. Cowburn, “Direct Measurement of 15 N Chemical Shift Anisotropy in Solution,” *Journal of the American Chemical Society*, vol. 120, no. 42, pp. 10947–10952, Oct. 1998.

[7] D. Fushman, N. Tjandra, and D. Cowburn, “to direct determination of protein dynamics from 15N NMR relaxation at multiple fields, independent of variable 15N chemical shift anisotropy and chemical exchange,” *Journal of the American Chemical Society*, vol. 121, pp. 8577–8582, 1999.

[8] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, “Atomic-level characterization of the structural dynamics of proteins.,” *Science (New York, N.Y.)*, vol. 330, no. 6002, pp. 341–6, Oct. 2010.

[9] I. Q. Phan, J. Boyd, and I. D. Campbell, “Dynamic studies of a fibronectin type I module pair at three frequencies: Anisotropic modelling and direct determination of conformational exchange.,” *Journal of biomolecular NMR*, vol. 8, no. 4, pp. 369–78, Dec. 1996.

[10] G. Lipari and A. Szabo, “Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results,” *Journal of the American Chemical Society*, vol. 2, no. 1, 1982.

[11] G. Lipari and A. Szabo, “Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity,” *Journal of the American Chemical Society*, vol. 104, no. 17, pp. 4546–4559, Aug. 1982.

[12] G. G. Dodson, D. P. Lane, and C. S. Verma, “Molecular simulations of protein dynamics: new windows on mechanisms in biology.,” *EMBO reports*, vol. 9, no. 2, pp. 144–50, Feb. 2008.

[13] M. Karplus and J. Kuriyan, “Molecular dynamics and protein function.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 19, pp. 6679–85, May 2005.

[14] M. Tirion, “Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis.,” *Physical review letters*, vol. 77, no. 9, pp. 1905–1908, Aug. 1996.

[15] E. Eyal, L.-W. Yang, and I. Bahar, “Anisotropic network model: systematic evaluation and a new web interface.,” *Bioinformatics (Oxford, England)*, vol. 22, no. 21, pp. 2619–27, Nov. 2006.

[16] S. M. Hollup, G. Salensminde, and N. Reuter, “WEBnm@: a web application for normal mode analyses of proteins.,” *BMC bioinformatics*, vol. 6, no. 1, p. 52, Jan. 2005.

[17] W. Zheng and S. Doniach, “A comparative study of motor-protein motions by using a simple elastic-network model.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 23, pp. 13253–8, Nov. 2003.

[18] E. J. Bertaccini, J. R. Trudell, and E. Lindahl, “Normal-mode analysis of the glycine alpha1 receptor by three separate methods.,” *Journal of chemical information and modeling*, vol. 47, no. 4, pp. 1572–9, Jan. 2007.

[19] A. Bakan, L. M. Meireles, and I. Bahar, “ProDy: protein dynamics inferred from theory and experiments.,” *Bioinformatics (Oxford, England)*, vol. 27, no. 11, pp. 1575–7, Jun. 2011.

[20] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation,” *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, Mar. 2008.

[21] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, “CHARMM: the biomolecular simulation program.,” *Journal of computational chemistry*, vol. 30, no. 10, pp. 1545–614, Jul. 2009.

[22] A. Hoeben, B. Landuyt, M. S. Highley, H. Wildiers, A. T. Van Oosterom, and E. A. De Bruijn, “Vascular endothelial growth factor and angiogenesis.,” *Pharmacological reviews*, vol. 56, no. 4, pp. 549–80, Dec. 2004.

[23] N. Ferrara, “Role of vascular endothelial growth factor in regulation of physiological angiogenesis.,” *American journal of physiology. Cell physiology*, vol. 280, no. 6, pp. C1358–66, Jun. 2001.

[24] N. Ferrara and R. S. Kerbel, “Angiogenesis as a therapeutic target.,” *Nature*, vol. 438, no. 7070, pp. 967–74, Dec. 2005.

[25] A. R. Cardones and L. L. Banez, “VEGF inhibitors in cancer therapy.,” *Current pharmaceutical design*, vol. 12, no. 3, pp. 387–94, Jan. 2006.

[26] S. Iyer and K. R. Acharya, “Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines.,” *The FEBS journal*, vol. 278, no. 22, pp. 4304–22, Nov. 2011.

[27] W. J. Fairbrother, M. A. Champe, H. W. Christinger, B. A. Keyt, and M. A. Starovasnik, “1H, 13C, and 15N backbone assignment and secondary structure of the receptor-binding domain of vascular endothelial growth factor.,” *Protein science : a publication of the Protein Society*, vol. 6, no. 10, pp. 2250–60, Oct. 1997.

[28] Y. a Muller, B. Li, H. W. Christinger, J. a Wells, B. C. Cunningham, and a M. de Vos, “Vascular endothelial growth factor: crystal structure and functional mapping of the kinase domain receptor binding site.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 14, pp. 7192–7, Jul. 1997.

[29] Y. a Muller, H. W. Christinger, B. a Keyt, and a M. de Vos, “The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 A resolution: multiple copy

flexibility and receptor binding.," *Structure (London, England : 1993)*, vol. 5, no. 10, pp. 1325–38, Oct. 1997.

[30] B. A. C. Horta, J. J. V. Cirino, and R. B. de Alencastro, "On the structure, interactions, and dynamics of bound VEGF.," *Journal of molecular graphics & modelling*, vol. 26, no. 7, pp. 1091–103, Apr. 2008.

[31] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay, "Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation," *Biochemistry*, vol. 33, no. 19, pp. 5984–6003, May 1994.

[32] R Core Team, "R: A Language and Environment for Statistical Computing." Vienna, Austria, 2012.

[33] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †," *The Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6474–6487, Jul. 2001.

[34] R. Anandakrishnan, B. Aguilar, and A. V Onufriev, "H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations.," *Nucleic acids research*, vol. 40, no. Web Server issue, pp. W537–41, Jul. 2012.

[35] T. Reddy and J. K. Rainey, "Interpretation of biomolecular NMR spin relaxation parameters.," *Biochemistry and cell biology = Biochimie et biologie cellulaire*, vol. 88, no. 2, pp. 131–42, Apr. 2010.

[36] J. García de la Torre, M. L. Huertas, and B. Carrasco, "HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations.," *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 147, no. 1, pp. 138–46, Nov. 2000.

[37] J. G. Kempf, J.-Y. Jung, C. Ragain, N. S. Sampson, and J. P. Loria, "Dynamic requirements for a functional protein hinge.," *Journal of molecular biology*, vol. 368, no. 1, pp. 131–49, Apr. 2007.

[38] L. E. Kay, D. A. Torchia, and A. Bax, "Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease," *Biochemistry*, vol. 28, no. 23, pp. 8972–8979, Nov. 1989.

[39] M. A. S. Hass and J. J. Led, "Evaluation of two simplified 15N-NMR methods for determining micros-ms dynamics of proteins.," *Magnetic resonance in chemistry : MRC*, vol. 44, no. 8, pp. 761–9, Aug. 2006.

[40] J. M. Kneller, M. Lu, and C. Bracken, "An effective method for the discrimination of motional anisotropy and chemical exchange.," *Journal of the American Chemical Society*, vol. 124, no. 9, pp. 1852–3, Mar. 2002.

[41] D. Fushman, S. Cahill, and D. Cowburn, "The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration.," *Journal of molecular biology*, vol. 266, no. 1, pp. 173–94, Feb. 1997.

[42] O. Walker, R. Varadan, and D. Fushman, "Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from (15)N relaxation data using

computer program ROTDIF.," *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 168, no. 2, pp. 336–45, Jun. 2004.

[43] J. L. Goodman, M. D. Pagel, and M. J. Stone, "Relationships between protein structure and dynamics from a database of NMR-derived backbone order parameters.," *Journal of molecular biology*, vol. 295, no. 4, pp. 963–78, Jan. 2000.

[44] U. Holzgrabe, I. Wawer, and B. Diehl, *NMR Spectroscopy in Drug Development and Analysis*. John Wiley & Sons, 1999, p. 299.

6.

Summary and discussion

6.1. Qualitative assessment of molecular recognition energetics in gas phase.

(*Michael Goldflam expressed all VEGF samples that were used in this work and performed chemical shift perturbation NMR experiments***)**

In the article 1 we addressed the question of how gas-phase stabilities of non-covalent interactions are related to the solution binding affinities. As we discussed in the introduction, it is possible to maintain non-covalent complexes in gas-phase in near-native state. Their stabilities in gas phase can be assessed qualitatively using various dissociation techniques. However, gas phase generally do not correlate positively with the affinities in solution. The systematical investigation of this phenomenon is important for understanding of the role of solvent in molecular recognition and for assessment of capabilities of mass-spectrometry as a technique for rational ligand design and drug development.

As a model system for our studies we used VEGF – a validated drug target associated with pathological angiogenesis. We expressed 11-109 fragment of VEGF-A, that retains the activity of full-length VEGF-A_{121} at the same time featuring higher solubility than wild type protein.

Phage display-derived cyclic peptide was picked as a ligand. This peptide is known to bind VEGF with dissociation constant K_D of about 1 μM. The area of binding of this peptide largely overlaps with the interaction interface of VEGF with its receptors (Figure 6.1).

In this work the peptide is referred to as **P-wt**. It is a 2340 Da 19-amino acid soluble peptide, stabilized with C5-C15 disulfide bridge. It is disordered in solution, however upon binding to VEGF it adopts well-defined conformation with α-helix at C-terminus (W13-E17). The hydrophobic residues are mainly positioned towards the surface of the protein, while polar residues point out, suggesting that protein-ligand interactions are formed at least partly thanks to the hydrophobic effect.

Since there is no well-established way to accurately calculate gas-phase stabilities of the non-covalent complexes, the only reliable information that can be obtained from the series of non-covalent complexes in the gas phase is their stabilities relatively to each other. It is much more straightforward to do so if all the complexes have the same molecular weight.

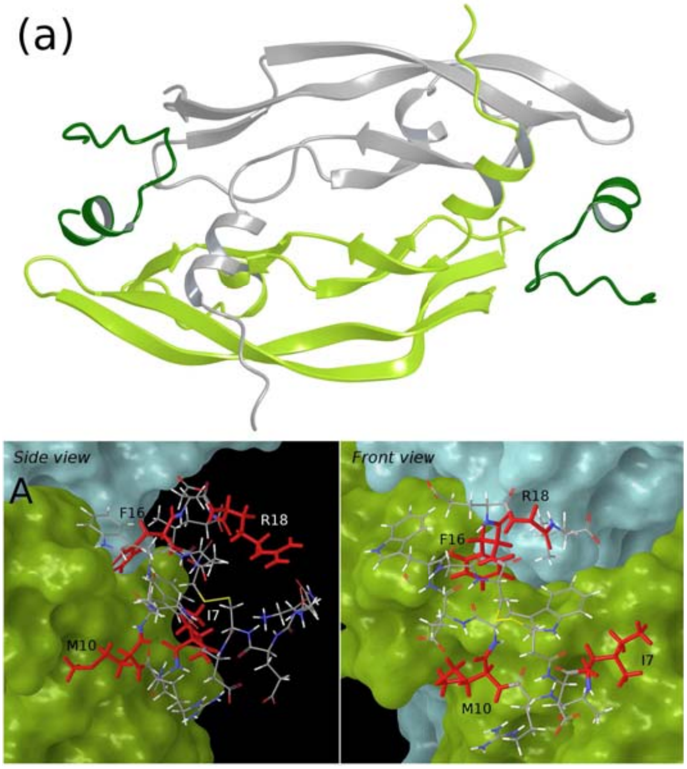


Figure 6.1. (a) NMR structure of VEGF with bound peptide **P-wt**; (b) close-up of bound peptide **P-wt** with the residues that underwent modifications highlighted in red.

Thus we introduced point modifications at different residues of **P-wt** peptide substituting the **L** amino acid with the **D** analog. The peptides are listed in the Table 6.1. The **P-wt** and 4 analogs were synthesized using solid-phase peptide synthesis. The conditions were optimized for the manual synthesis following standard Fmoc-chemistry strategy, and final synthesis of all five peptides were performed using automatic microwave peptide synthesizer.

Affinities of the peptides to VEGF in solution were assessed using chemical shift perturbation NMR technique. VEGF was titrated with each of five peptides and ^{1}H-^{15}N-HSQC spectra were acquired for each titration point (Figure 6.2a). For each ligand, four peaks with the strongest displacement of chemical shift $\Delta\delta$ were used. $\Delta\delta$ was calculated for each point using equation 6.1. Titration curves of $\Delta\delta$ plotted against the ligand concentration *[L]* were fitted to the equation 6.2 and the K_D was found for each peptide (see Figure 2b for example). The values of K_D are listed in the Table 2, and free energies of binding, calculated from K_D are plotted on the figure 6.3. It can be seen that point mutations at different point caused different effect on the affinity of the ligand to VEGF. As expected, substitution of hydrophobic Met10 and Phe16 that were directly involved in hydrophobic interactions with VEGF with **D** analogs resulted in dramatic decrease in binding affinity (3 and 2 orders of magnitude respectively). On the other hand, change of polar Arg18, that was pointing out of the binding interface, have much smaller effect on binding. Surprisingly, substitution of the Ile7 residue that according to

Table 6.1. list of cyclic peptides used as ligands for VEGF. Lower case bold letters represent **D**-amino acid residues.

Name	***Sequence***
P-wt	GGNECDIARMWEWECFERL
P-7i	GGNECD**i**ARMWEWECFERL
P-18r	GGNECDIARMWEWECFE**r**L
P-16f	GGNECDIARMWEWEC**f**ERL
P-10m	GGNECDIAR**m**WEWECFERL

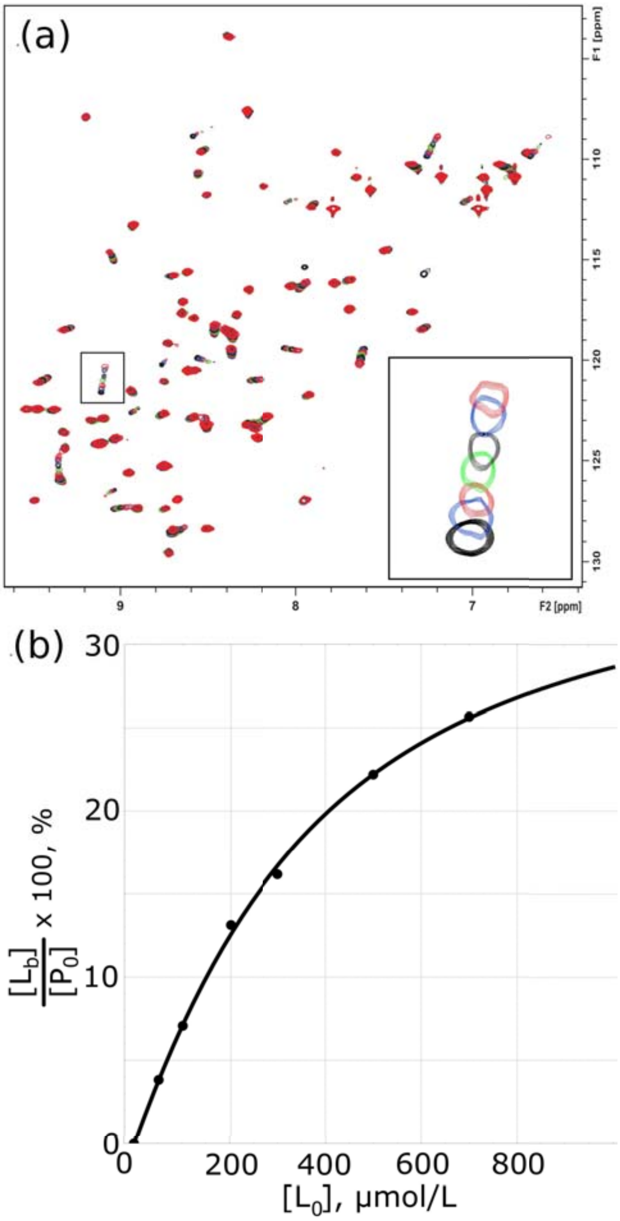


Figure 6.2. (a) ^{15}N-^{1}H HSQC spectra of a 100μM VEGF titrated with **P-7i** ligand; on the inset: zoom of Lys48 shifts; (b) Fit of Lys48 chemical shift perturbation.

$$\Delta\delta_{NH} = \sqrt{\Delta\delta_H^2 + \left(\frac{\Delta\delta_N}{5}\right)^2} \tag{6.1}$$

$$\Delta\delta_{NH} = k\frac{K_D + [L_0] + 2[P_0] - \sqrt{(K_D + [L_0])^2 + 4[P_0](K_D - [L_0] + [P_0])}}{2[P_0]} \tag{6.2}$$

where $\Delta\delta_N$, $\Delta\delta_H$ and $\Delta\delta_{NH}$ are chemical shift perturbations in ^{15}N, ^{1}H plane and 2-D chemical shift perturbation, respectively, k is the coefficient of proportionality needed for the fitting purposes, $[L_0]$ and $[P_0]$ are the initial concentrations of ligand and protein respectively. For more details see the paper.

the NMR structure is not involved directly in contact with VEGF resulted in quite dramatic change in the K_D - nearly equal to that of Phe16. That suggests that non-polar Ile7 takes part in hydrophobic interactions with VEGF.

In order to rank five ligands according to their gas-phase affinity to VEGF we performed native CID ToF MS of five complexes. Experimental conditions were optimized that allowed detection of all five complexes. PL_2^{+10} peak (m/z 2821) was isolated, and the CID of it was performed. (Here and further in the text P, PL and PL_2 refers to protein alone and complex of protein with one and two ligands respectively; superscript indicates the charge state). Collision voltage was varied in the range that was sufficient to record full breakdown curve for each complex with 2V step. In the Figure 6.4a gradual dissociation of PL_2^{+10} is depicted. The primary pathway of dissociation can be represented as $PL_2^{+10} \rightarrow PL^{+8} + L^{+2} \rightarrow P^{+6} + 2L^{+2}$, indicating that dissociating ligand has two protons, most probably, on Arg9 and Arg18. Minor dissociation pathway are discussed in the article 4.

Volumes of all the peaks were extracted from the spectra and breakdown curves corresponding

Table 6.2. list of peptide ligands with corresponding values of dissociation constants and collision voltages.

Name	***K_D, μM***	***V_{C50}, V***
P-wt	*1.02 ± 0.18*	*24.93 ± 0.11*
P-18r	*3.50 ± 0.60*	*27.64 ± 0.20*
P-7i	*252 ± 77*	*25.17 ± 0.14*
P-16f	*313 ± 53*	*28.00 ± 0.18*
P-10m	*1810 ± 147*	*34.31 ± 0.10*

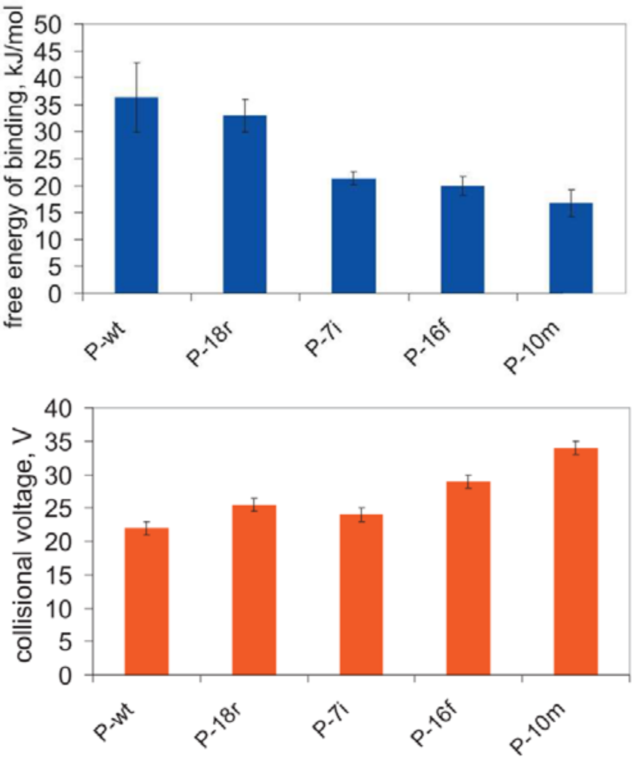
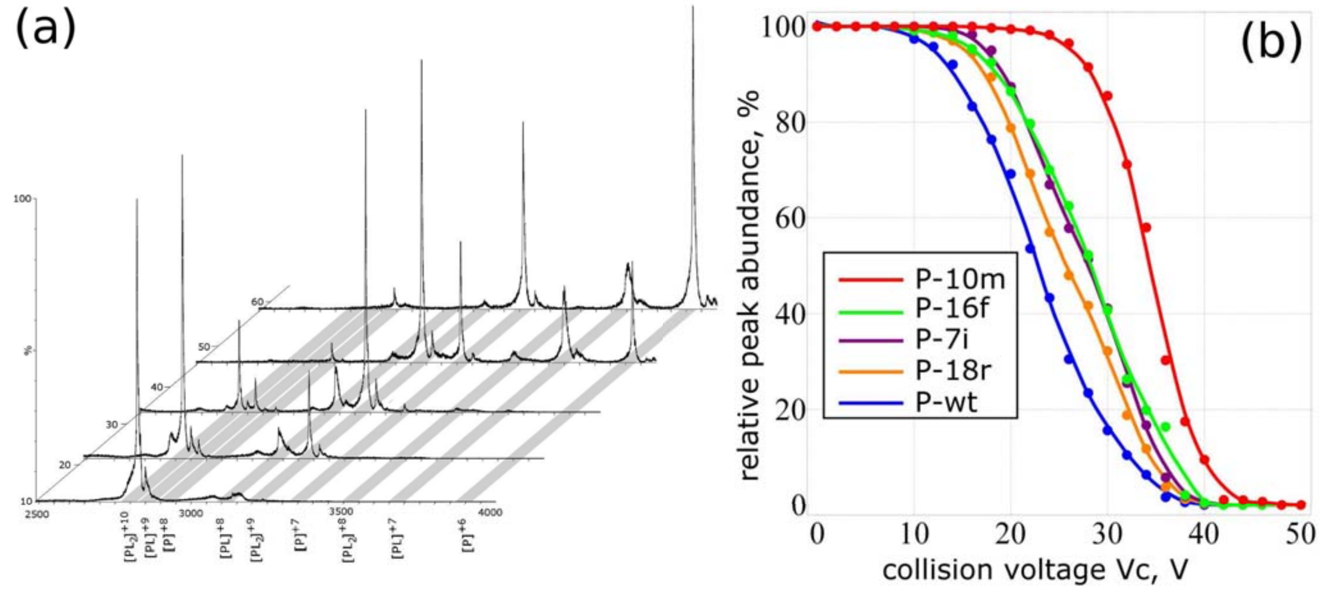


Figure 6.3. top: free energies of binding of 5 peptides to VEGF; bottom: collisional voltages needed to dissociate 50% of the non-covalent complex of VEGF with 5 peptides in gas phase.



Tigure 6.4. (a) Dissociation of VEGF PL_2^{+10} complex ion. Different spectra represent different collision voltages, plotted along y axis; (b) dissociation curves of complexes of VEGF with 5 peptides, corresponding to the $PL_2^{+10} \rightarrow PL^{+8} + L^{+2}$ dissociation pathway.

to the major $PL_2^{+10} \rightarrow PL^{+8} + L^{+2}$ dissociation pathway were constructed from them (Figure 6.4b). It can be seen that different ligands require different collision voltage to dissociate. For example, at V_C = 25V approximately half of all the VEGF-**P-wt** complex is dissociated, whereas the dissociation of VEGF-**P-10m** complex has barely started. The V_{C50} values (voltages needed for the dissociation of 50% of the complex) were calculated by fitting each breakdown curve to the cumulative distribution function of Gaussian distribution (see SI of the article 2 for more details). The values of V_{C50} are listed in Table 6.2 and plotted in Figure 6.3. It is clear that in gas phase as well as in solution five peptides show different affinity to VEGF. However, rankings of peptides in solution and in gas phase are reversed. That is, the wild type peptide **P-wt** that has the strongest affinity to VEGF in solution forms the weakest complex with VEGF in gas phase. On the other hand, the **P-10m** peptide that in solution has millimolar affinity to VEGF turns out to be the strongest binder in gas phase. This result is not very surprising taking into account considerations that were made in introduction and in the beginning of this section. But the underlying mechanism that leads to inverted affinities in solution and gas phase requires further investigation.

On a qualitative level one can construct the following theory. VEGF-**P-wt** binding is governed mainly by hydrophobic interactions, and its sequence is optimized by phage display to let it adopt the conformation that most efficiently exposes hydrophobic side-chains to the binding interface. The stronger perturbation we introduce to the hydrophobic surface by mutation, the weaker the hydrophobic interactions become and the worse is the binding in solution. On the other hand, the conformation of **P-wt** is amphipathic, and its structure is optimized to keep its charged residues as far from the hydrophobic patch on a surface of the protein as possible. Distortion of this structure is likely to allow polar side-chains to come in the vicinity of the protein binding interface. This has little effect in solution, when polar residues are screened by the solvation shell, but in gas phase in the absence of the solvent it can cause substantial gains in binding energy, effectively inverting the trend.

In order to get some insight in the thermodynamics of VEGF-peptide interactions we performed isothermal titration calorimetry (ITC) experiments on two non-covalent complexes – VEGF-**P-wt** and VEGF-**P-7i**. Unfortunately, it was impossible to study other three complexes by ITC because the concentrations required to detect the heat emitted upon binding of weaker ligands exceeded the solubilities of peptides. ITC is powerful biophysical technique that can yield accurate binding constant and, more importantly for our study, allows to separate entropic and enthalpic component of free energy of binding. Results of ITC (Figure 6.5) partly support this theory, revealing that the enthalpy of ligand binding to VEGF increased by 17 kJ/mol upon mutation of one residue. However, it revealed, somewhat unexpectedly, that binding of both ligands to VEGF are driven by the enthalpy, while entropy plays against them. This is not what one expects when the main component of interactions is hydrophobic. This seemingly contradictory result can be explained if we remember that **P-wt** and is disordered when free in solution. Binding to VEGF requires it to severely restrict the movement of its internal rotors, that results in a large entropic penalty, that cannot be balanced even by entropic gain from hydrophobic interactions with VEGF. In case of **P-18r** the picture is probably very similar, however its structure, distorted by the mutation, allows for less hydrophobic contacts with the protein, resulting in even stronger overall unfavorable entropy. As for the origin of favorable enthalpic component – careful examination of the structure of the VEGF-**P-wt** complex reveals that there is a possibility of formation of four hydrogen bonds: Phe16$_{\text{P-wt}}$ - Gln89$_{\text{VEGF1}}$, Trp11$_{\text{P-wt}}$ - Tyr21$_{\text{VEGF2}}$, Ile7$_{\text{P-wt}}$ - Asn62$_{\text{VEGF2}}$, and Arg9$_{\text{P-wt}}$ - Asn62$_{\text{VEGF2}}$ (here subscripts indicate that the residue belongs to the peptide, or to the first or second chain of VEGF). The energy from the hydrogen bonds can explain partly the strong enthalpy of binding, however they are not enough to explain all of it. The origins of the rest of enthalpy are not clear. In general, this behavior is an example a known phenomenon of entropy-enthalpy compensation.

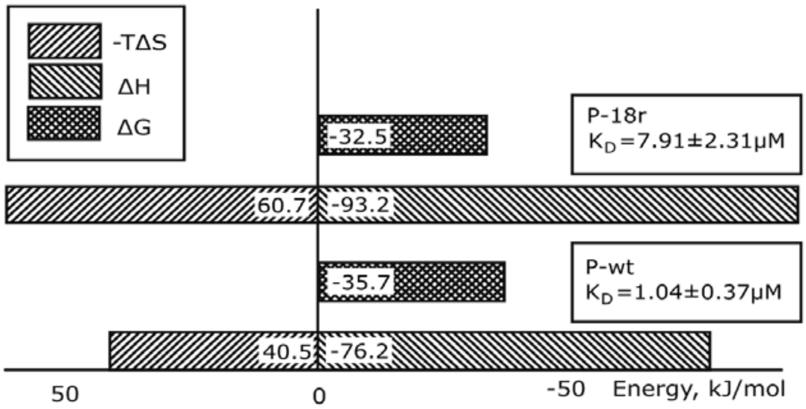


Figure 6.5. Free energy, enthalpy and entropic component of binding of two peptides (**P-wt** and **P-18r**) to VEGF.

In conclusion, we perform the study of an interesting model system that shows inversed stability ranking in solution and gas phase. This result suggests that at today level of understanding of the mechanism of non-covalent interactions their energy in gas phase cannot be taken as an estimate of solution binding energy, neither quantitatively nor qualitatively. On the other hand, it shows a potential of mass-spectrometry to provide insight into the nature of non-covalent interactions. However, a method that would allow accurate calculation of gas-phase binding energies would be highly beneficial for such investigations. This question is addressed in the next section.

6.2. Combined experimental/computational approach to gas-phase binding energy determination.

In the article 1 we showed that mass-spectrometry has a potency to provide additional information about the mechanism of non-covalent interactions. We also mentioned that this approach suffers from the lack of a good reliable method to estimate the internal energy that is deposited in the ion prior to fragmentation. In the article 2 we describe a method that allows modeling of the collisional excitation/relaxation of large molecular ions during CID and provides means to compute the internal energy that they gain under arbitrary experimental conditions.

Principles of CID fragmentation technique were discussed in the section 1.2.2 of the introduction. One of the key components necessary for the accurate description of energy transfer during collisional activation is an adequate collisional model. Exact description of inelastic ion-molecule collision requires taking into account large number of parameters, including shape of the ion, long-range ion-molecule interactions, distribution of incidental angles, probability of excitation of the gas molecule and a number of others. Collisional model that takes into account all these factors would be overcomplicated, impractical, difficult to develop and computationally expensive to apply. In order to construct a practical model, one needs to account for parameters that are critical for the particular case, and neglect those that has little effect on it.

The collisional model that is described in the article 2 was tailored to be used for CID of large molecular and complex ions in a multiple collision CID regime, typical for CID of non-covalent complexes. It is a statistical semi-empirical model, that is derived from basic physical principle of detailed balance and then parametrized for the specific system under study based on the experimental data. The detailed description and mathematics can be found in the supporting information. Here I will briefly describe the collisional model, Monte Carlo engine that is used to simulate collisions with gas molecules and finally the method to extract activation energy E_A of unimolecular dissociation. I will also describe two case studies to which the method was applied – short peptide and set of non-covalent complexes, described in the article 1.

In the present method the ion-molecule collision is treated as a statistical process, characterized by the probability density function $P(E,K_{COM},\Delta E)$, where E is internal energy of the ion, K_{COM} is the kinetic energy of ion-molecule system in the center of mass frame and ΔE is a change of internal energy of the ion, or energy step. This function defines the probability for the system with internal energy E in a collision with gas molecule with center of mass kinetic energy

K_{COM} to undergo the change of internal energy of the size ΔE. Note that ΔE can be positive (collisional excitation) or negative (collisional cooling). If the $P(E,K_{COM},\Delta E)$ is defined for a given ion on a three-dimensional grid of E, K_{COM} and ΔE, then it can be used to model its collisions and the evolution of internal energy. In the approach described in the article 2 $P(E,K_{COM},\Delta E)$ is calculated as follows. First, it is expressed as a product:

$$P(E, K_{COM}, \Delta E) = C(E, K_{COM}) f(K_{COM}, \Delta E) \tag{6.1}$$

where $f(K_{COM},\Delta E)$ is a probability for the system to undergo the energy change of the size ΔE in a collision characterized by COM kinetic energy K_{COM}, and $C(E,K_{COM})$ is a normalization coefficient. $f(K_{COM},\Delta E)$ is then defined arbitrarily; in order to compute $C(E,K_{COM})$, an integral equation is constructed utilizing two conditions: detailed balance principle and the normalization condition. This equation is solved numerically using a custom-built approach. For the details please refer to the supporting information for the article 2.

In order to model the ion-molecule collisions we use a Monte Carlo approach. The scheme of Monte Carlo algorithm is shown on Figure 6.6A. On each step the ion with velocity v_I collides with gas molecule with velocity v_G, that is generated randomly from Maxwell-Boltzmann distribution under experimental temperature. To model the event of collision, random number R ($0<R<1$) is generated and compared with pre-computed probability of negative energy step (F_{DS} on the scheme). If $R > F_{DS}$, the positive energy step is triggered (ion heating); otherwise – negative energy step (ion cooling). After the sign of the energy step ΔE is determined, its size is computed. The ion velocity after collision is calculated from the laws of energy and momentum conservation. More details can be found in supporting information for the article 2.

This collisional model is based on general principles and does not suffer from unrealistic assumptions. In order to be used for treatment of real-life experimental data it has to be applied in a correct environment. The motivation for the development of this approach was better understanding of experimental data, described in article 1, and that data was acquired on a Waters SYNAPT G1 instrument. CID in this instrument takes place in traveling wave ion guide (TWIG), hence it was necessary to model the ion-molecule collisions in the presence of alternating electric field of traveling wave. The principle of operation of TWIG is shown schematically on Figures 6.6B,C. Ions are propelled along the cell axis by means of the “waves” of electric potential, that are generated by the pairs of stacked ring electrodes. On the Figure 6.6B “H” stands for “high voltage” and “L” – for “low voltage”. When the voltage is applied to a pair of electrodes (“H” state) electric field in front of the “wave” accelerates ions in this area. After the period of time determined by the “wave velocity” experimental parameter the signal is switched to the next pair. In order to successfully pass through the TWIG ions have to acquire enough speed during their residence in the acceleration region. Otherwise they “roll over” the wave and are decelerated in the rear part of

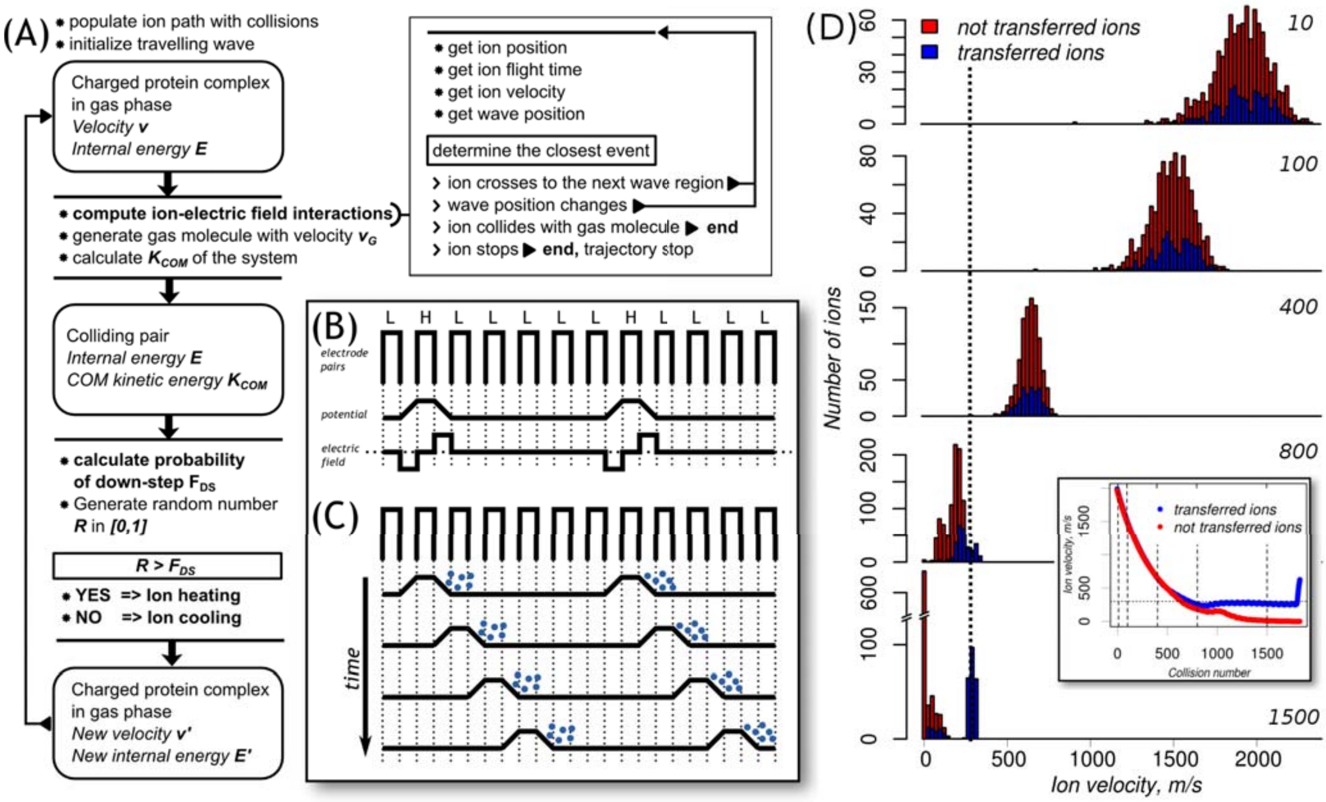


Figure 6. (A) principal scheme of the Monte Carlo engine; (B) schematic representation of the geometry of part of traveling wave ion guide, with shape of electric potential and electric field; (C) schematic representation of the ion transport in traveling wave ion guide; (D) velocity distribution of ions in traveling wave ion guide after different number of collisions. Blue bars correspond to ions that successfully reached the exit of the cell; red bars - to ions that stopped before that. Dotted line shows the traveling wave velocity. On the inset: average ion velocity plotted against number of collisions. Dashed lines show points at which histograms are plotted.

the wave. In the real instrument this causes ions to come in tight packs with narrow distributions of kinetic energies and times of arrival to the end of the cell.

The algorithm to model ion trajectories in TWIG is described in details in supporting information for the article 2. In short, it analyzes the velocity, position and residence time of the ion in collision cell on one side and position and operating time of the wave relatively to the electrodes on the other side (Figure 6.6A, right part). Based on this information algorithm chooses the next event among four possibilities: (1) ion crosses the position of the next electrode and passes on to the region between next two electrodes; (2) the position of the wave changes; (3) ion-molecular collision takes place; (4) ion velocity is zero or negative. In the cases 1 and 2 the ion position, velocity and residence time are updated and the next check is made. In case 3 the Monte-Carlo step is triggered, and in addition to position, velocity and residence time, ion internal energy is changed. In case 4 the trajectory is terminated.

Performance of the TWIG simulation algorithm with 1000 ions is illustrated on the Figure 6.6D. Plot shown on the inset depicts the evolution of average velocity for the ions that successfully reached

the exit of the cell (blue) and those that stopped before (red). Velocity distributions at the points marked by the dashed lines are shown on the main plot. It can be seen that at first ions rapidly lose their speed in collisions and distributions of “successful” and “unsuccessful” ions are similar. When the average velocity of ions comes close to the wave velocity (marked by the dashed line), “successful” ions stop decelerating, while “unsuccessful” keep losing speed in collisions and eventually come to a complete stop. The velocity of “successful” ions is narrowly distributed around the wave velocity, indicating that they are carried by the wave until the exit of the cell. Simulation algorithm represents the real-life behavior of ion in TWIG, in spite of the simplified one-dimensional electric potential configuration.

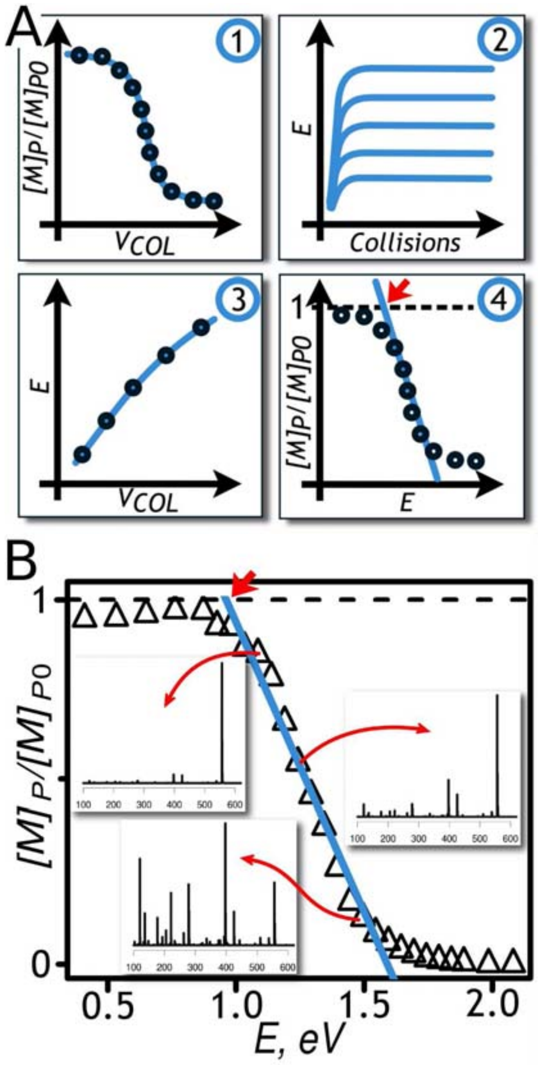


Figure 6.7. (A) Scheme of experimental data treatment workflow; (B) results of modeling of leu-enkephalin CID.

Now that both collisional model and the mechanism to simulate ion behavior in the TWIG ion guide are available, experimental data can be processed and E_A of unimolecular dissociation can be extracted. The workflow is shown on figure 6.7A. First, the percentage fraction of total ion population, that is not fragmented ($[M]_P/[M]_{P0}$, where $[M]_{P0}$ is initial parent ion population and $[M]_P$ is parent ion population at current collision voltage V_{COL}), is plotted against the collisional voltage (1). Next, after parameter optimization, a set of simulations is performed. Input collision voltages for these simulations should be evenly distributed along the range of experimental V_{COL} values, giving energies that are deposited in the ion at each voltage (2). The correlation between collision voltage and internal energy, deposited in the ion, is established using the results of the simulations (3), and the “real” breakdown curve is constructed. The “real” breakdown curve shows the fraction of not fragmented parent ion against the internal energy, deposited in the ion. In order to obtain the value of E_A from the breakdown curve, one needs to extrapolate its linear region until the interception with $[M]_P/[M]_{P0} = 1$ line. Internal energy at the interception level is an activation energy of unimolecular dissociation E_A.

Validation of this method is not a trivial task because the reliable quantitative information of gas phase stabilities of non-covalent complexes is scarce. In addition to that it is difficult to repeat the experiment for the same interacting partners at exactly the same conditions as authors used.

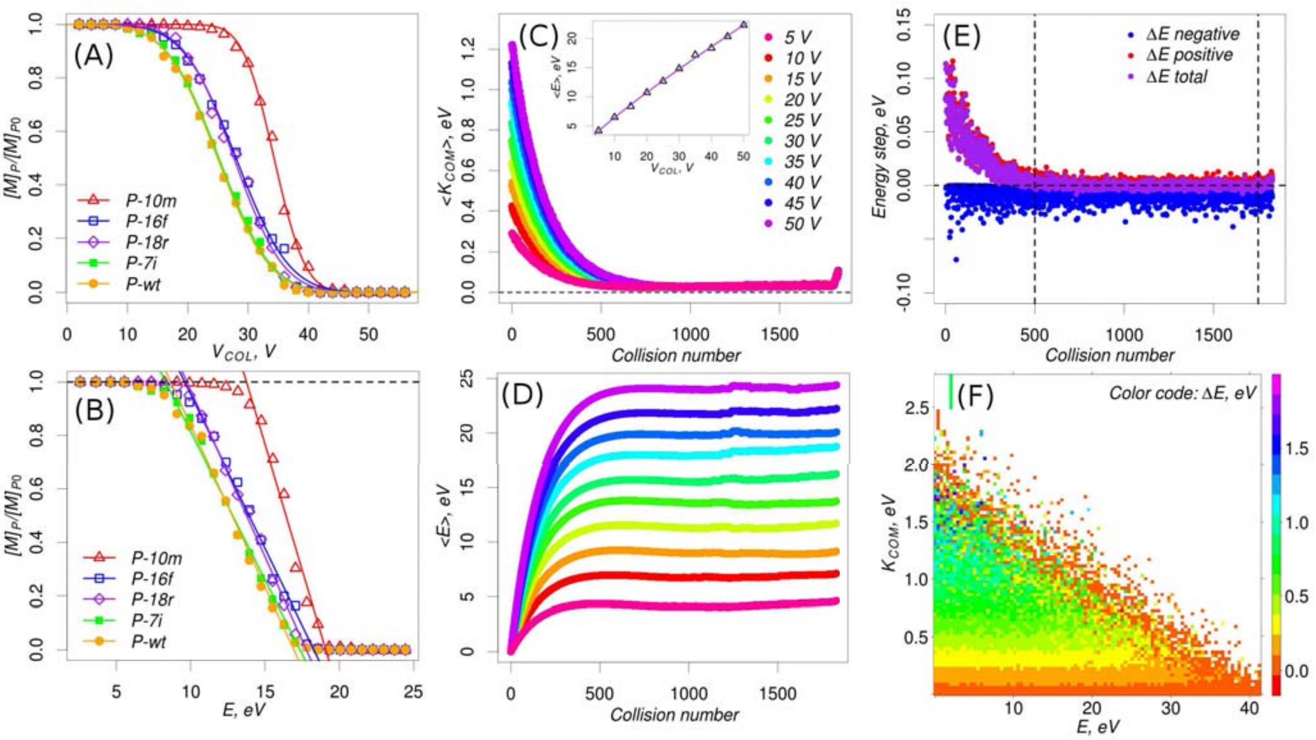


Figure 6.8. Treatment of 5 non-covalent complexes of VEGF with peptides. (A) Experimental collision voltage-dependent breakdown curves (points) and fitted curves (lines). (B) Internal energy-dependent breakdown curves (points) and fitting of their linear segments (lines). (C) Average COM kinetic energy of ion-molecule system, plotted against number of collisions. (D) Average internal energy of the ion, plotted against number of collisions. (E) Average positive, negative and total energy steps, plotted against number of collisions. (F) Level plot of energy steps, generated during ten simulation runs, plotted on a grid of COM kinetic energies and internal energies.

On the other hand, the model should perform equally well for the modeling of CID of covalent bonds of smaller molecules. However, they have to be large enough to fulfill $m_I >> m_G$ condition. Peptides appear to be good candidates, and we decided to validate the model using leu-enkephalin – 556.28Da 5 amino-acid neuropeptide with the sequence YGGFL. It has served as a model for many mass-spectrometry based experiments and an accurate value of E_A is published for it. CID of leu-enkephalin was performed on SYNAPT G1 instrument and the experimental data was used as an input for the simulations following the workflow described above. The final result is shown on Fugure 6.7B. The value of activation energy calculated by the method was 1.01 ± 0.05 eV, which is quite close to the published value (1.14 ± 0.05 eV). The discrepancy most probably comes from the fact that for smaller molecules the assumption that collisional cross-section is independent of COM kinetic energy of the system is less valid. This error is expected to be smaller for larger ions.

After validation with leu-enkephalin the model was applied to the set of non-covalent complexes between VEGF and 5 peptides described in the article 1. The results are shown on Figure 6.8. After construction of experimental breakdown curves the Monte Carlo algorithm has to be properly optimized. More specifically, the parameter that is responsible for ion cooling has to be set so that the system remained in a steady state through the entire trajectory (see article 2 SI for more details). The result of the optimization is shown on Figure 6.8E. The plot of average energy steps and positive and negative energy steps separately shows that after approximately 500 collisions the average energy step starts to be close to zero. This indicates that the system is in the steady state and Monte Carlo algorithm is parametrized correctly.

Figures 6.8C,D depict evolution of average values of COM kinetic energy and internal energy of the ion. Ions start with different K_{COM} that is defined by the input collisional voltage. They gradually lose their kinetic energy in collisions, and part of it is transferred into the internal energy. In each simulation the internal energy curve reaches saturation when ion loses its initial kinetic energy and its velocity is dictated by the traveling wave. The saturation level depends on the initial kinetic energy, and hence on the collision voltage. The energy steps, generated in this simulation are shown on Figure 6.8F in the form of a level plot. Clearly the size of energy step positively correlates with COM kinetic energy of colliding system. At the same time the correlation with internal energy is negative, because the probability of excitation decreases with the increase of internal energy. The final results are plotted on the Figure 6.8B, and the corresponding values of E_A for five complexes are listed in the Table 6.3. The activation energies calculated for **P-wt** and **P-7i** are the same within the error, which is not surprising taking into account that input collision voltages are very close. The same applies to peptides **P-18r** and **P-16f**. The highest activation energy corresponds to the **P-10m**. It is caused by the high input collision voltage and the steeper slope of its breakdown curve. The difference between activation energies of **P-wt** and **P-18r** is about 11%, which is not too far from the difference between enthalpies of binding, reported for the same two peptides in the article 1 (18%). However, the absolute value of this difference (9.74 eV – 8.64 eV = 1.1 eV ≈ 106 kJ/mol) is much higher than the difference between the enthalpies of binding (93.2 kJ/mol – 76.2kJ/mol = 17 kJ/mol). This in not surprising considering that in the aqueous media the effective dielectric constant varies from 4 to 80 and the energy of electrostatic interactions are scaled down by this factor relatively to the values in vacuum.

Table 6.3**.** List of peptides, used as ligands to VEGF with V_{C50} and values of calculated activation energies

Name	***V_{C50}, V***	***E_A, eV***
P-wt	24.93 ± 0.11	8.64 ± 0.63
P-7i	25.17 ± 0.14	8.31 ± 0.61
P-18r	27.64 ± 0.20	9.74 ± 0.71
P-16f	28.00 ± 0.18	9.58 ± 0.70
P-10m	34.31 ± 0.10	13.73 ± 1.00

The method presented in the article 2 allows modeling collisional activation of ions at wide range of conditions. Although it can theoretically be applied to any system, it is designed to perform best with large molecular and complex ions. It makes the method useful for study of non-covalent interactions in gas phase. The program is organized so that modules responsible for the treatment of ion-molecule collisions and modeling of ion trajectories are independent, making it easy to apply the method to a large variety of collision cells that are on the market today.

Overall, the work presented in the articles 1 and 2 provides an insight at the state of the art in the field of non-covalent interactions in gas phase. It is clear that there are still more questions then answers in this area. In the article 1 we formulate the problem, that can be summarized as follows: “How non-covalent interactions in gas phase are different from non-covalent interactions in solution and what can we learn from this difference?”. We present a good model system to study this phenomena and carry out a preliminary qualitative assessment of it in solution and gas phase. In the article 2 we develop an approach that provides means to solve some of these questions, namely, how much energy does ion acquire by collisions before dissociation. But there are still many questions to answer in this context before non-covalent interactions in gas-phase are completely understood, and this will be the subject of a future work.

6.3. Application of structural native mass-spectrometry to the study of allosteric interactions between GroEL molecular chaperonin and ATP.

(Ranit Gruber and Liat Shimon expressed GroEL samples that were used in this work; Prof. Amnon Horovitz performed the final analysis of allosteric regulations of GroEL by ATP)

In the article 3 we present an example of how unique features of mass-spectrometry can help to answer relevant biological questions. In particular, the ability of native mass-spectrometry to maintain native-like state of non-covalent complexes in combination with its high resolution allows direct detection of relative populations of different binding states, making it possible to study cooperativity and allostery of binding in complex systems.

The object of investigations in this work is GroEL, molecular chaperonin found in *e. coly* (Figure 6.9A), that is necessary for correct folding of some proteins in the cell and also assists in re-folding of misfolded proteins. The chaperone-assisted protein folding mechanism is highly conserved across species. The analog of GroEL in eukaryotes, Hsp60 shares many structural features with GroEL making a deciphering of its mechanism highly desirable for understanding of cellular functions. In addition to that it could shed light in thermodynamics of protein folding.

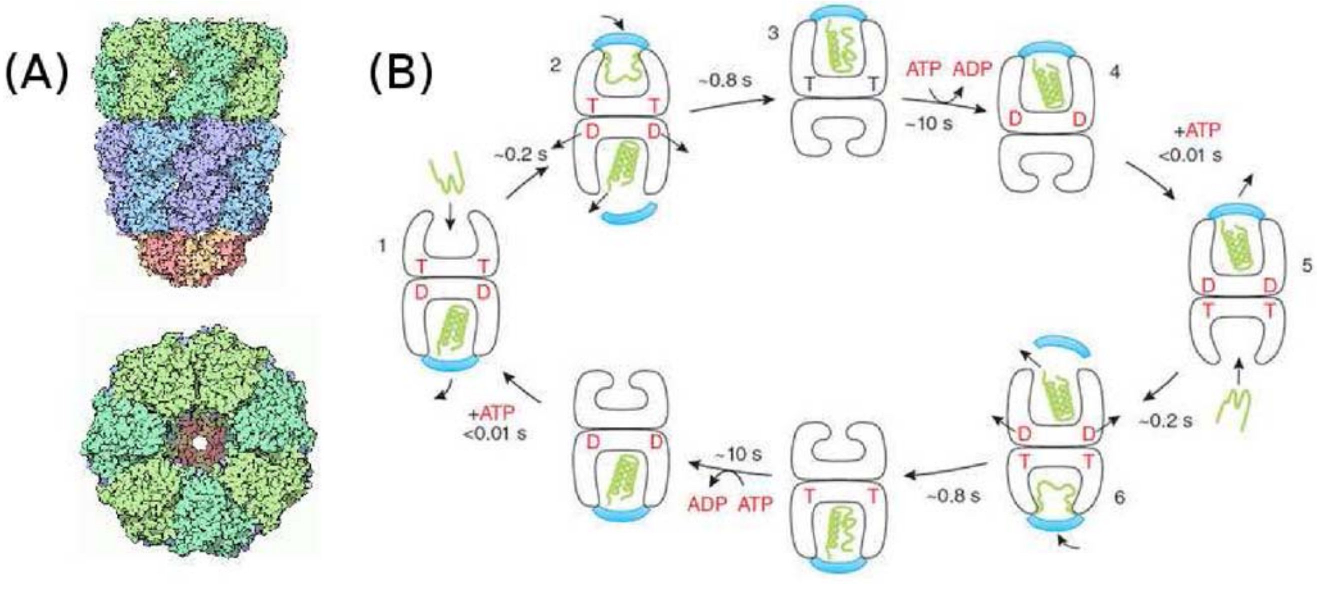


Figure 6.9. (A) Structure of GroEL (green,light blue and purple) and helper protein GroES (yellow, pink); (B) full cycle of function of GroEL (reprinted from [1]).

[1] A. L. Horwich, “Protein folding in the cell: an inside story.,” *Nature medicine*, vol. 17, no. 10, pp. 1211–6, Oct. 2011.

The cycle of function of GroEL is pictured on Figure 6.9B. As it can be seen, each ring of GroEL changes its conformation prior to and after binding the substrate. This conformational change is mediated by the binding and hydrolysis ATP. Moreover, it was shown previously using non-active mutants of GroEL that hydrolysis is not necessary for the conformational change – binding of ATP alone is enough. This suggests that study of different bound forms of GroEL by mass-spectrometry in a conditions where ATP hydrolysis is suppressed can provide information of real life behavior of GroEL.

The typical buffer that is used for native mass-spectrometry experiments is $C_2H_3O_2NH_4$ (ammonium acetate, in this text referred to as AmAc). This is a volatile buffer with weak buffering capacity at physiological pH, that is typically used in concentrations from 10mM and up to 5M.

After optimization of conditions and initial ESI-ToF MS experiments with GroEL in 1M AmAc buffer the following two weaknesses were revealed. First, the resolution that is achieved with AmAc was not sufficient to resolve two adjacent ATP bound states. The effective resolution in mass-spectrometry experiment is affected by the protonation state of the ion. Increase in the number of protons increase the value in denominator of measured quantity (m/z), effectively bringing two peaks closer together in the spectrum. Usual strategy to address this issue is to add charge-reducing agents, for example, imidazole. Spectra of GroEL tetradecamer in 1M AmAc with and without 10 mM imidazole is shown on Figure 6.10. Adding the charge reducing agents is undesirable because it can reduce the quality of the spectra and cause unwanted interactions with the system under study.

Second problem, more fundamental for the success of the planned research, arises from the fact that hydrolysis of ATP by GroEL is facilitated by K^+ ions. In order to obtain well resolved spectrum of ATP bound states it is necessary to suppress the ATP hydrolysis. It was planned to achieve this by putting GroEL in the potassium-free environment. However, it was found that NH_4^+ ions of AmAc can to some extent mimic potassium ions. As a result, ATP is getting hydrolyzed completely by GroEL on the time scale of the experiment. This finding rendered AmAc buffer useless for our experiments.

In the pursuit of a suitable buffer solution we looked for volatile water-soluble compounds that have some buffer activity in the physiological pH

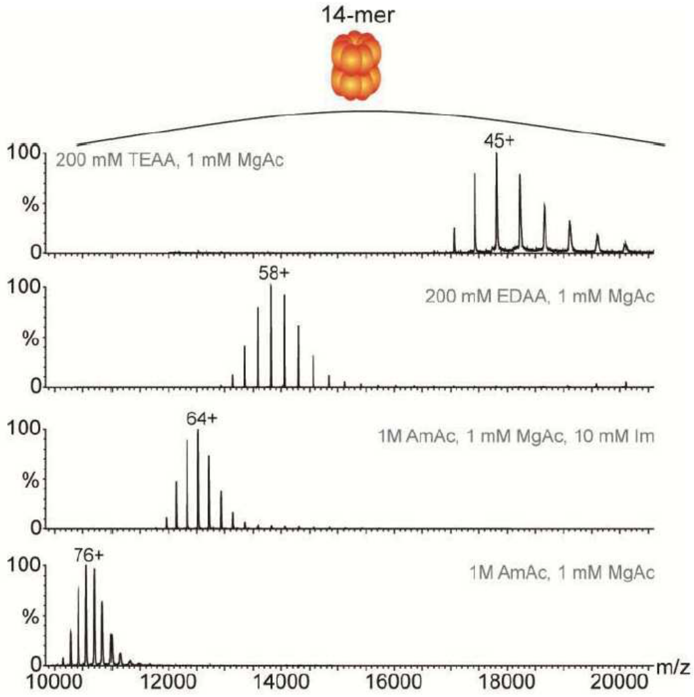


Figure 10. ESI-ToF MS spectra of GroEL in different volatile buffers. From bottom to top: 1M AmAc, 1M

AmAc with 10mM imidazole, 200 mM EDAA, 200 mM TEAA. 1mM MgAc is added to each solution in order to provide Mg^{2+} ions that are necessary for proper function of GroEL.

region. Two candidates were tried. First, triethylammonium acetate ($(CH_3CH_2)_3NHC_2O_2H_3$, in this text referred to as TEAA) was reported to be used in native mass-spectrometry experiments. Second, ethylenediammonium acetate ($C_2H_4(NH_2)OCOCH_3$, here referred to as EDAA) to the best of our knowledge was never reported as a buffer for native mass-spectrometry. Spectra of GroEL tetradecamer in 200mM TEAA and 200mM EDAA are shown on the Figure 6.10. It can be seen that in both buffer solutions GroEL ions acquire less protons than in AmAc. TEAA shows the most dramatic effects – average charge state is almost two times smaller than in AmAc. This would make it ideal candidate for our experiments. However, the resolution that can be achieved in TEAA was rather poor. On the contrary, spectra acquired in EDAA showed excellent resolution, that was comparable to the one achieved in non-native conditions (with addition of 20% MeOH, for example). No ATP hydrolysis was detected in EDAA, so it was picked for the experiments.

1 μM GroEL in EDAA buffer (200mM EDAA, 1mM $Mg_2(CH_3CO_2)_2$, pH 7.0) was titrated with ATP at concentrations 0.469 μM, 0.938 μM, 1.25 μM, 1.532 μM, 1.875 μM, 2.5 μM, 5 μM, 6.125 μM, 7.5 μM, 10 μM, 12.5 μM, 15 μM, 17.5 μM, 20 μM, 25 μM, 30 μM, 35 μM, 40 μM, 50 μM, 60 μM and 100 μM. The fragment of the spectrum of GroEL with 5 μM of ATP is shown on Figure 6.11A. Multiple protonation states are available, making it possible to reduce statistical errors by using data from different charge states. Each peak was fitted to the set of Gaussian curves, and fitting error was minimized using PeakFit4 program (Systat Software Inc.). The smoothed peak of +58 charge state of GroEL with 5 μM of ATP and its reconstruction from fitted Gaussians are shown on Figures 6.11B and C respectively. Similar analysis was performed for three charge states at each concentration (Figure 6.11D).

To be able to correlate binding state populations, derived from ESI MS one has to consider the effects of non-specific binding. Non-specific binding occurs inevitably during electrospray ionization process when solvent evaporates from protein-containing droplets and concentration of non-volatile compounds increases for short period of time. In order to correct the data for the non-specific binding, we used method developed by L. Shimon et. al. [2]. The formula, presented in the article [2] contains an error. On top of that, it can be applied only to the cases when the peak corresponding to protein with no ligand bound is present in the spectrum, which is not always the case in the GroEL-ATP titration data. The suggested mathematical method was corrected (equation 6.2) and extended to the case in which the intensity of free protein peak is zero (equation 6.3). These formulas were used to remove the effects of non-specific binding from experimental results. The plot of fractional saturation Y of GroEL with ATP vs ATP concentration is shown of Figure 11E. Y is was calculated according to equation 6.4. Y reflects the overall level of saturation of protein

---

[2] L. Shimon, M. Sharon, and A. Horovitz, "A method for removing effects of nonspecific binding on the distribution of binding stoichiometries: application to mass spectroscopy data.," *Biophysical journal*, vol. 99, no. 5, pp. 1645–9, Sep. 2010.

that the experimental data can be described better with concerted MWC cooperativity model, rather binding sites with ATP. The acquired data allowed to treat each bound state individually and thus than sequential KNF model (see article 3 SI for more details). According to this model, protein exists in two distinct conformation: one with lower affinity to the ligand (T) and one with higher affinity (R). Switch between two states is initiated by binding of certain number of ligand molecules and takes place in a concerted manner (for more details see article 3 and its supporting info). The relative populations of T and R states at different bound states are shown on Figure 6.11F.

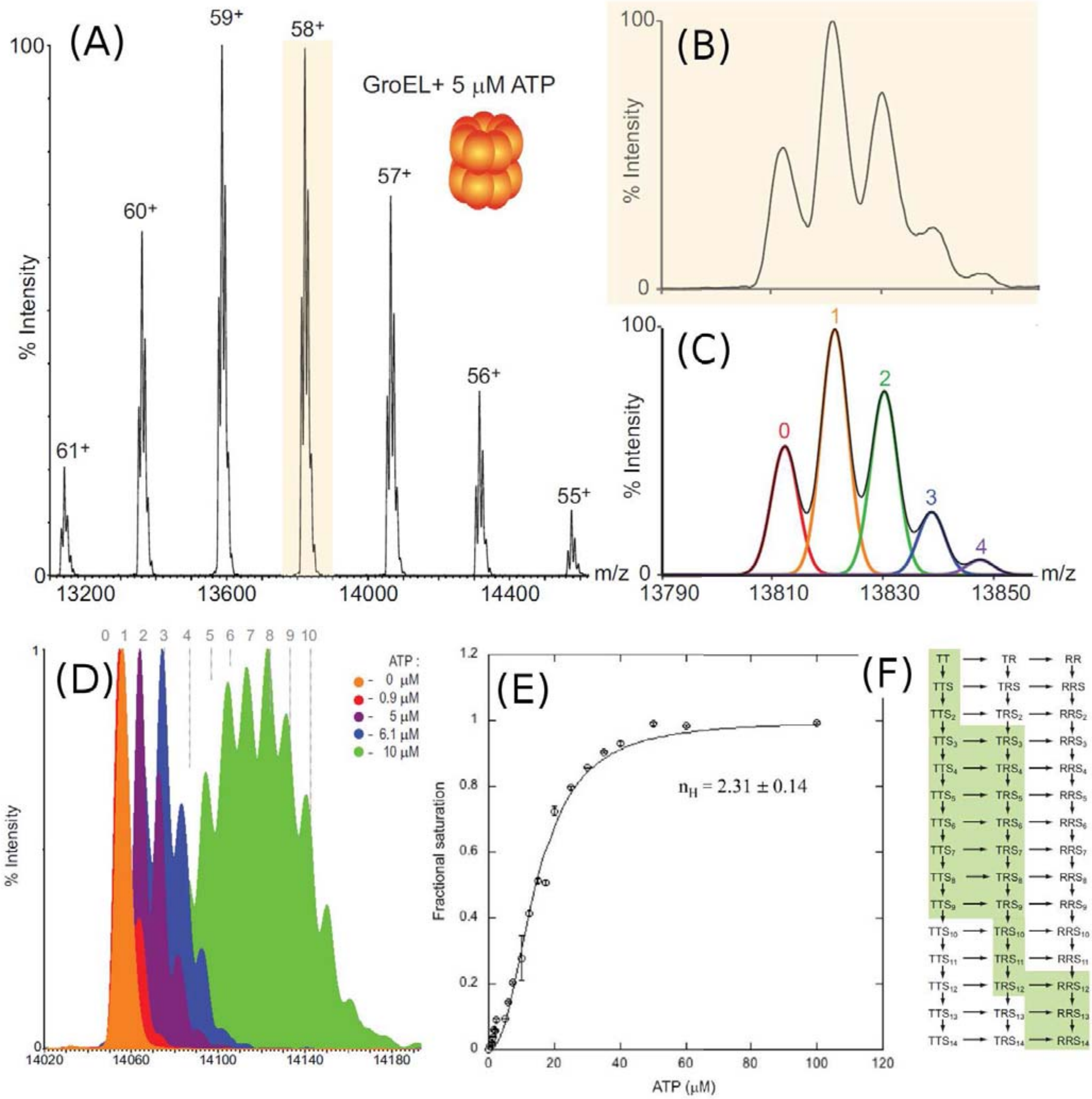


Figure 6.11. (A) Fragment of the ESI-ToF mass-spectrum of GroEL + 5 μM ATP in EDAA buffer; (B) zoom on +58 charge state; (C) reconstruction of +58 charge state from the best fit Gaussians; (D) superposition of +57 charge state spectra of solutions of GroEL with different concentration of ATP; (E) fractional saturation curve of GroEL–ATP binding (see text for more details); (F) scheme showing the of balance between of different conformations of two rings of GroEL depending of number of ligands bound; the most populated conformations are highlighted in green.

$$\frac{C_N}{[E]} = \left(\frac{I_N - I_{N-1}[S]}{I_0}\right)\left(\sum_{j=1}^{\alpha-N}(K_n[S])^{j-1}\right) \tag{6.2}$$

$$\frac{C_N}{[E]} = \left(\frac{I_N}{I_{N-1}} - K_n[S]\right)\left(\sum_{j=1}^{N}\left[[S]^{N-1}K_n^{j-1}\prod_{i=1}^{N-j}K_i\right]\right)\left(\sum_{j=1}^{\alpha-N}(K_n[S])^{j-1}\right) \tag{6.3}$$

$$Y = \frac{\sum_{i=1}^{M} iC_i}{M\sum_{i=1}^{M} C_i} \tag{6.4}$$

where C_N is a total population of protein species with N ligand molecules bound specifically, [E] and [S] are concentrations of enzyme in substrate respectively, I_N is the intensity of peak in the spectrum corresponding to N ATP molecules bound, Kn is a non-specific binding constant, K_i is a binding constant for *i-th* specifically bound ligand molecule and *α* is the maximum number of different bound states visible in the current spectrum, M is the total number of specific binding sites (14 in case of GroEL)

The conformations with highest population are highlighted in green. It can be seen how at first both rings exist in low-affinity conformation. It was shown previously that strong negative cooperativity exists between two rings of GroEL, suggesting that the majority of ATP molecules bind to one ring.

After 3 ATP molecules are bound, one ring starts to switch to the high-affinity conformation, that is associated with higher values of K_i. TT and TR states co-exist in the equilibrium when the first ring approaches saturation, and when 9 molecules of ATP are bound to GroEL population of TT conformation becomes negligibly small. Finally, when 12, 13 and 14 molecules are bound, second increase of K_i is detected, that is associated with switch of GroEL to the RR form. Thus it was possible to reconstruct a picture of MWC-type allosteric regulation of GroEL by ATP.

This work highlights the potential of native MS in structural biology thanks to its unique ability to determine the relative populations of co-existing species with different number of bound ligands. In this example we showed how it can serve to distinguish between two cooperativity mechanisms and construct a full picture of allosteric regulation of GroEL by ATP. Although this result is valuable by itself, it can be also seen as a proof of concept. There are many biological systems that would highly benefit from the similar investigations. In this context, minor questions that were solved in this work, namely, development of mathematical formalism for removal of the effects of non-specific binding and discovery of EDAA buffer should facilitate the research in this field.

6.4. Backbone dynamics of VEGF studied by the combination of NMR relaxation and normal mode analysis.

(Dr. Margarida Gairi performed the NMR relaxation experiments)

Articles 1-3 were dedicated to the study of PPI and PLI in solvent-free environment. In the article 4 I would like to present the work that is aimed at another factor, that plays crucial role in the mechanism of non-covalent interactions – conformational plasticity of proteins their ligands. As it was pointed out in the section 1.3 of introduction, proteins are highly dynamical systems, and we are very far from understanding how protein dynamics affects protein-protein interactions.

In this work we studied backbone flexibility of VEGF protein using NMR relaxation and normal mode analysis. The application of NMR to study backbone dynamics is discussed in details in the section 1.3.2 if introduction. Here I will summarize the results and discuss them.

Two sets of NMR relaxation parameters were acquired. Longitudinal relaxation time T_1, transverse relaxation time T_2 and ^{1}H-PN nuclear Overhauser effect (NOE) were obtained at 14.4 T (600MHz) and 18.1 T (800 MHz) static magnetic fields. The results was evaluated in order to check for the possible aggregation. The overall rotational correlation time τ_c, calculated from the T_1/T_2 ratio showed good agreement with the simulated value, confirming the absence of aggregation. The raw

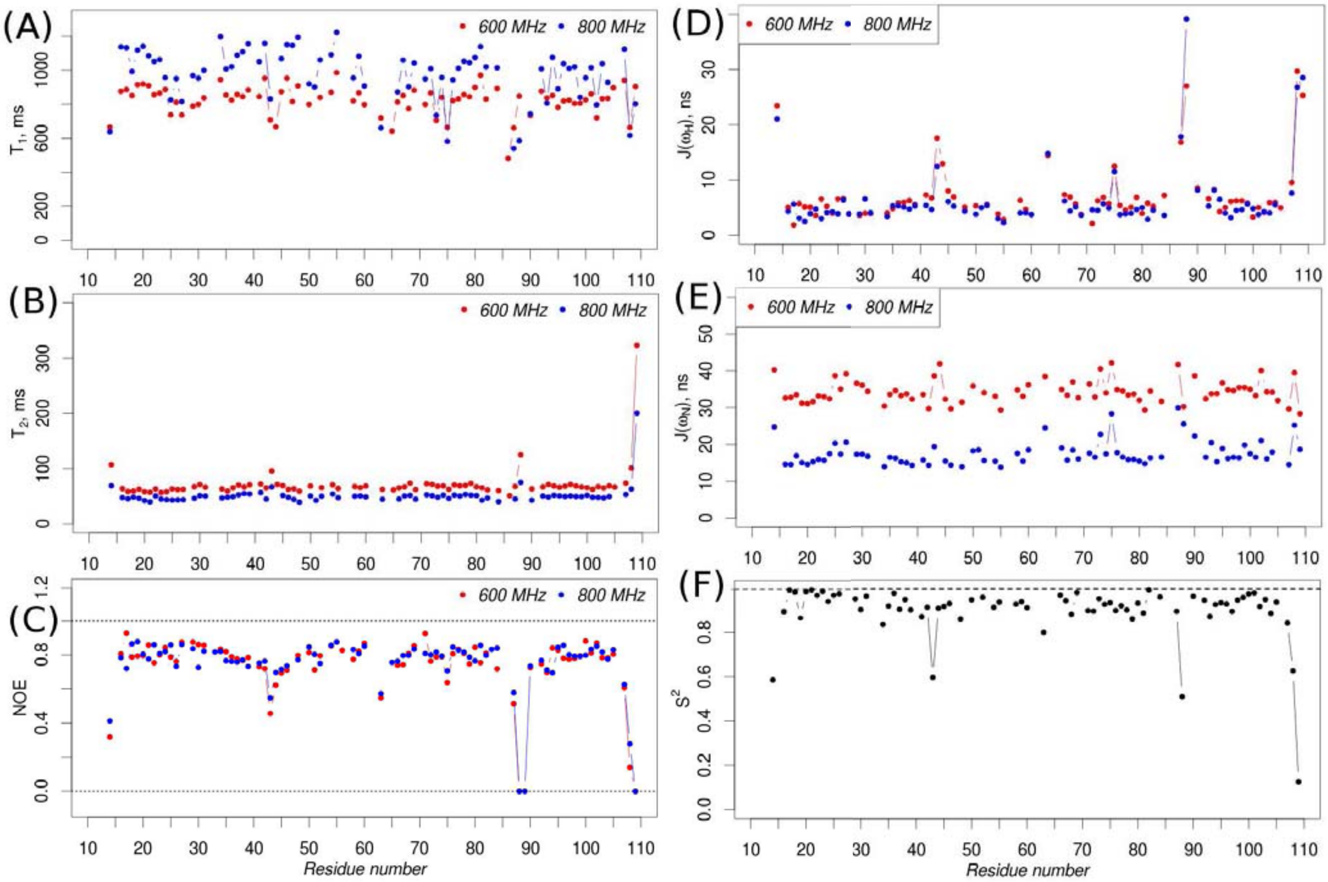


Figure 6.12. (A,B,C) NMR relaxation parameters, acquired a for VEGF at 45°C at two static magnetic fields: 14.4 T (600 MHz) and 18.1 T (800 MHz), T_1, T_2 and ^{1}H-^{15}N-NOE respectively; (D,E) spectral densities calculated from NMR relaxation parameters at ^{1}H and ^{15}N Larmour frequencies respectively; (F) generalized order parameter, derived from model-free analysis of relaxation data acquired at 14.1 T.

data was processed and analyzed by nmrPipe program. The results are shown on Figure 6.12 A,B,C.

There are some preliminary conclusions that can be drawn from the raw relaxation data. Lower values of T_2 can indicate that the corresponding residue is involved in motion on μs to ms time scale (chemical exchange), while elevated T_2 may point on the residues involved in motion at ps to ns time scale. Lower values of NOE are also very often associated with fast (ps to ns) dynamics. In case of VEGF, NOE and T2 clearly point at residues D63 and Q87-G89, situated on loop 1 and loop 2 respectively. In addition to that, NOE and T_2 also indicate possibility of ps-ns motion of I43 residue, located in an unstructured region of VEGF backbone. At the same time, no residues with T_2 significantly lower than average is found, hence no evidence of chemical exchange. The absence of chemical exchange is further confirmed by the plots of T_1/T_2 and $1/(T_1{*}T_2)$, that are specifically used to identify residues involved in motion on μs to ms time scale (for more details see article 4).

In order to get more comprehensive picture of VEGF backbone dynamics, we applied RSDM approach. RSDM allows to calculate residue-specific spectral density $J(\omega)$ at three frequencies: ω_H, ω_N and 0. Conceptually, spectral density mapping treatment can benefit greatly from the availability of data acquired at two distinct static magnetic fields (so called multi-field approach). The RSDM treatment of the data obtained at a single magnetic field suffers from numerous generalizations, two most harmful being (1) chemical exchange has zero effect on transverse relaxation time; and (2) the chemical shift anisotropy (CSA) of ^{15}N nuclei is independent of its chemical environment and can be considered constant for all residues. These assumptions do not affect $J(\omega_H)$, however correct value of CSA is needed for $J(\omega_N)$ and both CSA and chemical exchange rate constant R_{ex} are necessary for the accurate calculation of $J(0)$. In case of VEGF it seems that R_{ex} can indeed be considered zero, which makes multi-field approach even more attractive because it could allow for direct calculation of residue-specific CSA (see article 4 for more details).

Unfortunately, multi-field approach requires that both data sets were acquired at exactly the same conditions. The differences in concentration and buffer of two samples that were used for acquisition of two sets of relaxation parameters seem to be enough to render the approach ineffective for this case. Thus it was decided to apply standard single-field RSDM that, although less accurate, can still provide qualitative information about backbone flexibility. The value of CSA used for all residues across the protein was -160 ppm. The results are shown on Figures 6.12D,E. There are clearly four regions that exhibit higher flexibility than the rest of VEGF backbone. These are residues E44 (loop 1), D63 (unstructured region), Q87, G88 and H90 (loop 2), and E73 and N75 (β-sheet 5). Not surprisingly, it confirms preliminary conclusions made from T_2 and NOE.

The data was also treated using model-free formalism. The fitting was more successful for the 600 MHz dataset. The plot of squared generalized order parameter S^2 is shown on fig 1D. The average value of S^2 is 0.90 ± 0.13, confirming that VEGF is a very rigid protein. Residues that feature lower values of S^2 (I43, D63 and G88) are situated in previously identified flexible areas of VEGF backbone, confirming previous findings. The deeper insight into the backbone flexibility of VEGF is hampered by the fact that relaxation parameters are not available for about 20% of the backbone NH vectors. Normal mode analysis (NMA) is an attractive computational technique because it is computationally inexpensive and rather straightforward to apply. Nevertheless, if used correctly it can provide a complementary information to NMR relaxation about the dynamics of the protein. In this work we applied GROMACS 4.5.3 molecular dynamics simulation package to to perform all-atom force-field based NMA of VEGF. RMSD of backbone α-carbons was calculated for each of first 40 modes. The result is shown on Figure 6.13.

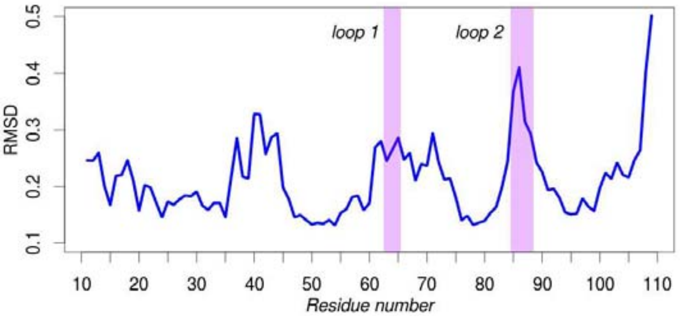


Figure 6.13. Mean scaled Cα RMSD derived from all-atom force-field based NMA. Two flexible loops are highlighted in pink.

It can be seen that four flexible areas, predicted by NMA, are largely overlapping with the ones predicted by NMR relaxation. This fact is better illustrated by Figure 6.14. Blue and red bars always point to the same regions of VEGF backbone. The discrepancies can sometimes be explained by the lack of sequence coverage by NMR. In other cases they probably arise from the fact that first normal modes take into account large overall motion of the protein. Thus the method can point out the flexible regions of the protein, but doesn’t reach single residue resolution.

Another important observation can be made from comparison of the flexible areas of VEGF with the zones that are involved in VEGF-receptor interactions (purple bars on Figure 6.14). As it can be seen, most of the resdues that are involved in VEGF-reciptor binding are situated either in or close to the flexible areas of VEGF backbone. This suggests that flexibility plays an important role in non-covalent interactions of VEGF with its two receptors.

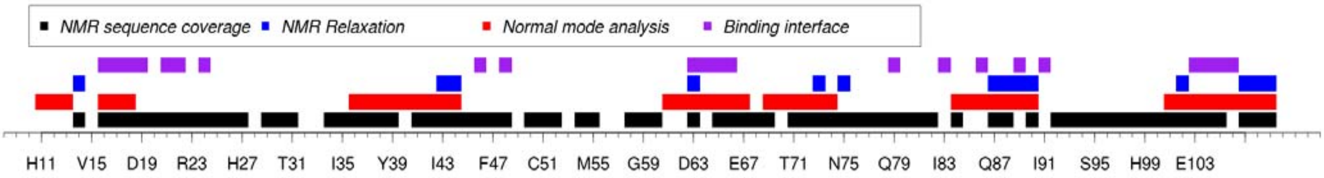


Figure 6.14. Comparison of flexible regions of VEGF backbone, predicted by NMR relaxation (blue) and NMA (red) with residues, involved in VEGF-receptor interactions (purple). Black bar indicates residues for which full set of three relaxation parameters was available at 600 MHz, 800 MHz or both.

6.5. Concluding remarks

Our current level of understanding of the “language” of non-covalent interactions between biological molecules does not allow us to “speak” freely on it, and the work presented in this thesis only scratched the surface of this difficult issue. The role that the solvent plays in protein-protein and protein-ligand interactions is far from being completely understood. Similarly, no universal strategy is developed to date that would fully account for the role of conformational plasticity in weak interactions between complex molecules. However, we showed that a combination of experimental and computational approaches can provide valuable information for both cases.

Examination of non-covalent complexes in gas-phase can shed some light on the mechanisms of interplay between the interacting interfaces and the aqueous media. We have shown that the behavior of non-covalent interactions changes dramatically upon removal of the aqueous media and discussed the possible source of these changes. They mainly arise from the fact that the strength of hydrophobic interactions greatly decreases in the absence of polar solvent, while all the interactions that are of electrostatic nature become much stronger. This changes result in a shift of thermodynamic equilibrium. However, thanks to the fact that on the time scale of MS experiment many non-covalent complexes preserve native-like contacts, comparison of the binding energies in solution and gas phase can provide interesting conclusions about the role of water. We showed here that combination of rigorous theoretical modeling and well-established CID-MS experimental approach can provide an insight into energetics of non-covalent interactions in gas phase. We have also shown that data, obtained from MS experiments on large protein complex, if corrected and treated appropriately, can provide unique quantitative information about the populations of different bound states, leading to important conclusions about the allosteric effects in protein-ligand interactions.

In addition to that we have demonstrated that combination of NMR relaxation experiments with normal mode analysis provides higher confidence in identification of flexible regions of protein backbone. In case of VEGF, we have discovered that the these flexible regions largely coincide with VEGF-receptor interaction interface, suggesting that flexibility is important for these interactions. This could facilitate design of new molecules able to modulate VEGF-receptor interactions and regulate VEGF-induced angiogenesis.

In this work we presented our attempts to tackle two major difficulties that complicate our understanding of non-covalent interactions. These difficulties hamper our ability to design specific binders from the scratch, avoiding long and laborious process of ligand optimization. This work, although far from finding solutions for these problems, represents certain advancement in understanding of the role of solvent and dynamics in non-covalent interactions and shows how multidisciplinary approaches can aid in finding solutions for complex questions.

7.

Conclusions

Conclusions:

Hence, in the view of the results presented, the main conclusions of the present thesis are the following:

- The strength of protein-ligand non-covalent interactions in gas-phase do not reflect the solution binding affinities neither quantitatively nor qualitatively, as shown on the example of non-covalent interactions of VEGF with five cyclic peptides.
- Analysis of non-covalent interactions by mass-spectrometry can provide valuable qualitative information about the role of solvent. However, for a quantitative assessment a reliable method to compute gas-phase stabilities of non-covalent complexes is required.
- The combined experimental/computational method reported in this thesis allows to model collisional activation of large molecular and complex ions during CID MS. The method is universal and can be applied to different systems with no theoretical upper limit for molecular weight. However, it is designed to perform better with large ions.
- The developed approach can also be applied to the different CID cells, found in commercial mass spectrometers. This has been illustrated by simulations of ion trajectories in the traveling-wave ion guide, used by SYNAPT G1 spectrometer from Waters.
- Native MS was applied to the study of allosteric interactions between GroEL and ATP. The scheme of balance between two conformations of GroEL was constructed for different number of bound ATP molecules. Strong evidence in favor of MWC cooperativity model was obtained.
- During this study a new buffer native MS-compatible buffer, ethylenediammonium acetate, was discovered. It allowed to achieve higher resolution than traditionally used buffers.
- The previously described mathematical approach for removal of the effects of non-specific binding from quantitative MS data was corrected and extended.
- Backbone dynamics of VEGF was studied combining NMR relaxation studies and normal mode analysis (NMA). It was shown that NMA can provide relevant information about protein flexibility and complement NMR relaxation studies.
- Four flexible zones were identified on the surface of VEGF. It was shown that those flexible zones, to a large extent, overlap with the interface of interactions between VEGF and its receptors.

8.

Breve resumen de la tesis

8.1 Resumen de los articulos

8.1.1 Articulo 1: Molecular recognition at protein surface in solution and gas phase: Five VEGF peptidic ligands show inverse affinity when studied by NMR and CID-MS.

Andrey Dyachenko, Michael Goldflam, Marta Vilaseca, Ernest Giralt,

Durante los últimos años, la técnica ESI-MS ha emergido como un instrumento de gran utilidad para el estudio de interacciones no covalentes. La espectrometría de MS se diferencia de otras estrategias biofísicas porque permite realizar estudios en el vacío, entorno que comporta cambios sustanciales en el comportamiento de las moléculas. En el presente trabajo se estudian las interacciones establecidas por VEGF con cinco ligandos peptídicos para los que se supone diferente afinidad. Con este objetivo, se sintetizaron un ligando peptídico [1], descrito previamente en la literatura, y reemplazando cada uno de los L-aminoácidos constitutivos por el correspondiente D-residuo, cuatro análogos diferentes. Las interacciones de VEGF con estos 5 ligandos fueron estudiadas por calorimetría de titulación isotérmica (ITC) y NMR, pudiéndose determinar mediante la aplicación de dichas técnicas las correspondientes constantes de disociación. La estabilidad en fase gas fue evaluada usando la técnica CID-MS.

El modelo de transferencia de energía, descrito en [2], fue adaptado para calcular la energía de interacción, y los ligandos peptídicos fueron ordenados según la afinidad mostrada con respecto VEGF tanto en solución como en fase gas. Los resultados muestran que la clasificación de los péptidos sigue el orden inverso en fase gas que en solución. En el presente artículo, se discute esta observación considerando que las diferentes interacciones no covalentes que estabilizan el complejo proteína-ligando evolucionan de manera muy diferente al eliminar el disolvente.

8.1.2 Articulo 2: Quantitative assessment of gas-phase stabilities of non-covalent complexes.

Andrey Dyachenko, Ernest Giralt

Las interacciones no covalentes presentes en las macromoléculas biológicas son esenciales para las funciones vitales de los organismos vivos. Durante los últimos años, mucho esfuerzo ha sido invertido en desarrollar nuevas técnicas que permitiesen estudiar los complejos proteicos no covalentes, tanto en fase gas como en solución. En oposición a una gran variedad de métodos en solución que ofrecían tanto información cualitativa como cuantitativa, las estrategias en fase de gas desarrolladas hasta el momento no permitían determinar la energía de estabilización de dichos complejos no covalentes en ausencia de disolvente. En el presente trabajo, describimos una innovadora estrategia basada experimentos de espectrometría de masas de disociación por colisión inducida, que permite determinar la energía de activación de la disociación unimolecular en fase gas.

Esta aproximación combina principios básicos de la física con la técnica de simulación de Monte Carlo, para modelar la excitación/relajación de iones en el momento de la colisión. El método es fácilmente modulable y puede ser aplicado a una gran variedad de células de colisión, incluyendo “travelling-wave ion cells”, encontradas en instrumentos modernos de tipo rayo. La estrategia diseñada establece las bases para la determinación precisa de estabilidades de unión de complejos no covalentes proteína-proteína y proteína-ligando en fase gas, que puedan ser estudiados mediante instrumentación moderna de espectrometría de masas.

8.1.3 Articulo 3: Allosteric mechanisms can be distinguished using structural mass spectrometry

Andrey Dyachenko, Ranit Gruber, Liat Shimon, Amnon Horovitz and Michal Sharon

Debido a la pequeña diferencia de masa entre apo-GroEL y sus derivados unidos a nucleótidos, las condiciones usadas en los experimentos de espectrometría de MS fueron optimizadas para alcanzar picos sumamente resueltos. Una serie de tampones volátiles compatibles con la técnica de espectrometría de masas y con notable capacidad para incrementar la precisión de la medida de masas, fueron estudiados y evaluados. Tampones como el acetato de amonio fueron desestimados debido a que los cationes NH_4^+ pueden desempeñar el mismo papel que los cationes K^+ en la hidrólisis del ATP promovida por GroEL.

En base a los experimentos realizados, descubrimos que el acetato de ethylenediammonio, un tampón que no había sido usado previamente en espectrometría de masas, promovía una resolución óptima de los picos, permitiendo de esta manera distinguir entre los complejos de GroEL y sus diferentes ligandes nucleotídicos. Las poblaciones relativas de las diferentes especies del complejo GroEL fueron determinadas por medición de las áreas de los picos asignados a cada uno de los derivados. En este punto, la contribución de la unión no específica entre ATP y GroEL fue eliminada mediante la aplicación de un tratamiento matemático desarrollado por nosotros, y que hace uso de los datos de espectrometría de masas para calcular dicha interacción. De este modo, se obtuvo la distribución real del complejo GroEL y de sus derivados unidos a ATP.

Considerando estas distribuciones de poblaciones, fue posible calcular los valores de las 14 posibles constantes de unión del ATP con el complejo GroEL. Usando estos valores, fuimos capaces de definir las diferentes vías de funcionalización del complejo GroEL con moléculas de ATP, y establecer el mecanismo allostérico de dicho complejo, el cual es crucial para la función que desempeña. Además, también pudo estudiarse la variación que experimenta el coeficiente de Hill, nH, y demostrar su aumento y luego su disminución en función de la concentración de ligando. En líneas generales, el trabajo descrito aquí demuestra que la aplicación estructural de la espectrometría de MS permite analizar la co-existencia entre una proteína libre y sus numerosos derivados resultantes de la unión de dicha proteína a un número variable de moléculas de ligando, permitiendo de esta manera distinguir entre diferentes modelos alostéricos.

8.1.4 Articulo 4: Backbone dynamics of vascular endothelial growth factor studied by NMR relaxation combined with normal mode analysis.

Andrey Dyachenko, Margarida Gairí, Ernest Giralt

En el presente estudio, se marcó con^{15}N el constructo más pequeño de VEGF que mantiene la misma afinidad por sus dos receptores, este es el constructo 11-109 de VEGF-A_{121}. Los expetrimentos de T_1, T_2 y ^{15}N {^{1}H} NOE fueron realizados bajo dos campos magnéticos diferentes, 600 MHz y 800 MHz. Las señales de, aproximadamente, el 75 % de los 99 residuos fueron obtenidas con la resolución suficiente para ser analizadas. “Reduced spectral density mapping” y “model-free analysis” fueron empleados para el estudio de los datos espectroscópicos. Disponer de los datos espectroscópicos a 600 MHz y a 800 MHz nos permitiría ganar precisión cuando se aplicó el “model-free fitting” y más fiabilidad en la zona de baja frecuencia cuando empleamos el “reduced spectral density mapping”.

Además de esto realizamos el análisis en modo normal de VEGF. El RMSD de los carbones alfa, calculado para los 40 primeros modos normales de VEGF, proporcionó información complementaria a los datos de relajación de RMN. Fue posible obtener información de aquellos residuos que habían sido inaccesibles para los experimentos de relajación. El análisis mostró que VEGF como un todo es una proteína muy rígida. Las regiones que muestran el mayor grado de movimiento interno en la escala de tiempo de pico- a nanosegundos, están situadas principalmente en la superficie de interación del VEGF con su receptor. Los residuos que exhiben movimiento interno elevado están situados en dos “loops”, ambos involucrados en la unión de VEGF con sus dos receptores. Estos datos contribuirán a la comprensión general de las interacciones que VEGF establece con sus receptores, y facilitarán el desarrollo de nuevos medicamentos antitumorales.

8.2 Introducción

8.2.1 Interacciónes no covalentes

Las biomoléculas de los organismos vivos realizan sus funciones principalmente a través de interacciones débiles reversibles entre ellas. La transducción de señal, la replicación de ADN/ARN, otros procesos enzimáticos y, virtualmente, cualquier otro proceso involucrado en las funciones vitales de cualquier organismo vivo (de las simples amebas, al complejo ser humano), requiere que las moléculas "hablen" entre ellas [3],[4]. Dicho lenguaje se basa en interacciones no covalentes.

Descifrar el mecanismo de cada interacción proteína-proteína o proteína-ligando no es una tarea fácil. Las interacciones proteína-proteína y proteína-ligando están caracterizadas principalmente por dos parámetros: la afinidad, o energía de unión entre las especies involucradas en la interacción; y la especificidad, o la capacidad de la proteína para reconocer a su compañero (o compañeros) e interaccionar selectivamente con él (o ellos). La capacidad de distinguir al compañero de interacción correcto de entre miles de otras moléculas requiere unos niveles extremos de especificidad extremos, que son alcanzados gracias a una modulación muy precisa de la superficie de interacción.

Además de la especificidad, cada interacción binaria se caracteriza por su afinidad. La energía liberada durante la unión de dos moléculas determina la fuerza de esta interacción. Esta energía va de 10 a 80 kJ/mol, que corresponden a afinidades de unión de 10^2 a 10^{14} M^{-1} [5],[6]. Las dos dificultades principales que impiden la descripción precisa de las interacciones no covalentes son la plasticidad de las proteínas y el efecto del disolvente.

El disolvente puede afectar la interacción de maneras muy diversas. Las moléculas de agua proporcionan un escudo electrostático alrededor de las especies cargadas (iones). Además de esto, cada grupo polar interacciona con el entorno dieléctrico (en este caso, el disolvente). La eliminación del disolvente resulta en una penalización energética sustancial en lo que a interacción entre proteínas o proteína-ligando se refiere [7].

La flexibilidad conformacional es una propiedad esencial de las grandes biomoléculas, y muchas de las funciones desempeñadas por proteínas se basan en su capacidad para cambiar de conformación en respuesta a un factor externo [8]. Geométricamente hablando, la presencia de flexibilidad en una proteína obstaculiza el diseño racional de medicamentos porque posibilita la existencia de un número muy elevado de conformaciones de dicha proteína. Por este motivo, cualquier información sobre la flexibilidad de una proteína es sumamente valiosa para la comprensión de PPI y PLI y para el diseño racional de medicamentos. Los capítulos 1-3 de la presente tesis versan sobre la solvatación, mientras que la flexibilidad se estudiará en el capítulo 4.

8.2.2. La espectrometría de masas

La espectrometría de masas es una técnica que permite la medición de la relación entre masa y carga de una especie iónica. Ésta puede aplicarse a una amplia gama de especies, desde los isótopos de diferentes elementos químicos [9], hasta virus enteros [10], pasando por pequeños compuestos químicos [11], biomoléculas mayores [12] y complejos no covalentes [13]. En todos estos casos el esquema de los experimentos de espectrometría de masas se mantiene esencialmente intacto (Figura 8.1).

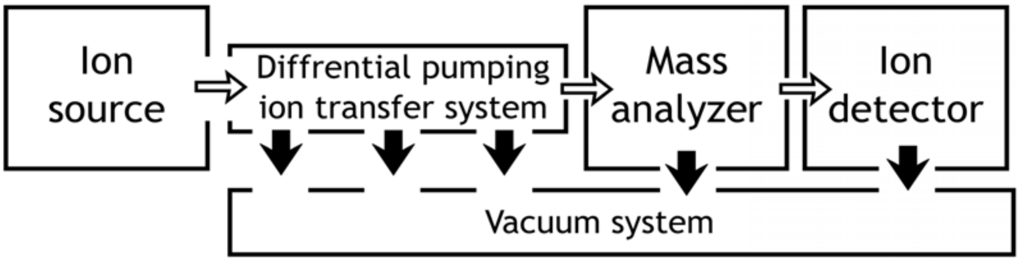


Figura 8.1. Esquema principal del espectrómetro de masas.

La "Collision induced dissociation" (CID) es una técnica de fragmentación comúnmente usada en espectrometría de masas. En un experimento típico de CID, los iones son acelerados mediante campos eléctricos e introducidos en una célula de colisións llena de gas inerte. El ión sufre colisiones con las moléculas de gas y su energía cinética se transforma en energía interna E. Cuando la energía interna del ión alcanza un cierto nivel, este comienza a disociar. Nos referiremos a esta energía como energía de activación de disociación unimolecular E_A [14]. La estrategia para determinar E_A para los iones más grandes como proteínas o complejos de proteínas no ha sido desarrollado hasta el momento.

La "native mass-spectrometry" es una técnica emergente que permite el estudio de proteínas y complejos proteicos enteros en la fase de gas, conservando su estado muy cercano al original (estado en solución) [15]. La importancia de dicha técnica en campos como la bioquímica [16], la biología estructural [17], y el desarrollo de nuevos medicamentos [18] está aumentando de manera sustancial durante los últimos años. El método aprovecha técnicas de iónización "suaves" (ESI, MALDI) que son capaces de iónizar y evaporar biomoléculas perturbando mínimamente su estructura nativa. La "native mass-espectrometry" abre posibilidades para las técnicas de "high-throughput screening" [18] así como para el estudio de complejos individuales no covalentes en fase gas [19–22].

Hay pruebas experimentales que sugieren que los complejos no covalentes mantienen su integridad en fase gas. Sin embargo, la energía de unión no covalente en fase gas no puede considerarse como una estimación de la energía de unión en solución [23][24]. Esto abre la posibilidad a investigar el papel del disolvente en interacciones no covalentes mediante la comparación de las energías de unión en solución y en fase gas. Para hacer un análisis cuantitativo, se requiere un método adecuado para el cálculo de energías de unión en fase gas. Este punto se trata en el capítulo dos de la presente tesis.

8.2.3 Papel de la dinámica en interacciones no covalentes y aplicación de la RMN en esta cuestión.

Es un hecho bien establecido que las grandes biomoléculas, en especial las proteínas, son sistemas sumamente dinámicos [25]. Este hecho tiene implicaciones inmediatas en los procesos de reconocimiento proteína-proteína y proteína-ligando [26][27]. Un gran número de aproximaciones computacionales están siendo desarrollados para representar estos procesos [28–30]. No obstante, los datos experimentales continúan representando la única fuente fiable de información sobre la dinámica de proteínas, y es por ello que son usados para parametrizar y validar los modelos computacionales.

Los experimentos de RMN de relajación se usan para caracterizar el movimiento molecular en una escala de tiempo que va de los picosegundos hasta los segundos [31]. Se miden tres parámetros diferentes: el tiempo de relajación longitudinal, el tiempo de relajación transversal y el efecto NOE (Nuclear Overhauser Effect). Los núcleos comúnmente más usados en los estudios de relajación de proteínas son ^{15}N y ^{13}C. Las dos estrategias más populares para analizar los parámetros de la relajación son “reduced spectral density mapping (RSDM)” [32] y “model-free formalism” [33][34].

8.2.4 El factor de crecimiento vascular endotelial

1.4. El factor de crecimiento vascular endotelial VEGF-A es un homodímero covalente antiparalelo globular sostenido gracias a dos puentes disulfuro intermoleculares y a interacciones hidrofóbicas. Cada monómero contiene, al menos, 6 puentes disulfuro intramoleculares, que forman puentes cistina muy estables [35]. Es una proteína altamente estructurada, cuya estructura secundaria contiene aproximadamente un 12 % de hélices y un 53 % de lámina β. Además, cada monómero presenta dos lazos, D63-G65 (lazo 1) y P85-G88 (lazo 2). Los residuos de los “loops” y de la hélice α situada en el extremo *N*-terminal (F17-Y25) están implicados en las interacciones de VEGF con sus receptores [36–38].

El factor de crecimiento vascular endotelial (VEGF) es un agente proangiogénico clave, responsable del crecimiento de la vascularidad nueva y de la dañada [39]. La sobre-expresión de VEGF-A está asociada con el crecimiento de tumores. Las células tumorales expresan grandes cantidades de VEGF induciendo la propagación de vasos sanguíneos en el tumor, facilitando de esta manera el suministro de sangre. La inhibición de la angiogénesis inducida por VEGF es una estrategia terapéutica aprobada, que se ha mostrado eficaz para suprimir la proliferación de tumores [40] y para aumentar su respuesta a la radioterapia [41]. Estas observaciones experimentales convierten a VEGF en un potente objetivo terapéutico, y promueven su estudio así como el desarrollo de nuevos inhibidores [42].

8.2.5 GroEL

GroEL es una proteína presente en una gran variedad de bacterias. Pertenece a la familia de las chaperonas y sus funciones principales son ayudar en el plegamiento de proteínas y reparar aquellas que se hayan plegado incorrectamente [43]. *E.coli* GroEL y su proteína ayudante GroES, son las chaperoninas más estudiadas hasta el momento [44]. Desde una perspectiva estructural, GroEL consiste en dos anillos de 7 miembros, apilados uno encima del otro, formando una estructura parecida a un barril. La masa de cada una de las 14 subunidades está alrededor de 57 kDa, dando un masa total del complejo alrededor de 800 kDa. El modo de acción de GroEL puede describirse, generalmente, de la manera siguiente [45].

El sustrato no plegado (o plegado incorrectamente) es encapsulado en la cavidad central de GroEL, que en ese punto queda obstruida por el heptámero GroES. Dentro de la cavidad se produce el plegamiento del sustrato. El mecanismo exacto de este proceso es aún desconocido. La encapsulación y la liberación del sustrato están asociados con un cambio conformacional de GroEL mediado por la unión y la hidrólisis alostérica de ATP.

8.2.6 Objetivos

En el contexto de los hechos presentados anteriormente, los tres objetivos principales de la presente tesis son:

1. Desarrollar una estrategia combinada (experimental/computacional) para la evaluación cuantitativa de la estabilidad de complejos no covalentes en fase gas.

2. Desarrollar una metodología para estudiar la regulación allostérica de grandes complejos proteicos usando espectrometría de masas estructural.

3. Estudiar la dinámica del esqueleto de VEGF utilizando una combinación de RMN de relajación y análisis de modos normales.

8.3 Resultados y discusión

8.3.1 Evaluación cualitativa de la energética de reconocimiento molecular en la fase gas.

(Michael Goldflam ha expresado todas las muestras de VEGF usadas en este trabajo y también ha llevado a cabo los experimentos de perturbación de desplazamiento químico (RMN))

En el articulo 1 realizamos la pregunta de cómo la estabilidad e los interacciones no covalentes en fase gas está relacionada con afinidades de unión en solución. La proteína VEGF fue usada como sistema modelo para nuestros estudios. El péptido cíclico obtenido por phage display, con afinidad para VEGF de 1 µM fue escogido cómo un ligando (Figura 8.2). Introdujimos modificaciones puntuales en diferentes residuos del péptido **P-wt** substituyendo el aminoácido **L** con el análogo **D**. Los 5 péptidos análogos están descritos en la Tabla 8.1.

Las afinidades de los péptidos respecto a VEGF en solución fueron evaluadas usando la técnica de RMN de perturbación del desclasamiento químico. Los valores de K_D están descritos en la Tabla 8.2, y las energías libres de unión calculadas a partir de valores de K_D están representadas en la Figura 8.3. Se puede observar que las mutaciones puntuales sobre las diferentes residuos causaron un efecto diferente sobre la afinidad del ligando a VEGF.

La estabilidad en fase gas de los 5 complejos no covalentes fue evaluada con CID ToF MS en las condiciones nativas. Las curvas de ruptura correspondientes al mayor camino de disociación ($PL_2^{+10} \rightarrow PL^{+8} + L^{+2}$), fueron realisadas a partir de cada experimento (Figura 8.4b). Se puede observar que las diferentes ligandos requieren diferente voltaje de colision para la disociación. Los valores de V_{C50} estan descritos en la Tabla 8.2 y representados en la Figura 8.3.

Es evidente que en fase gaseosa, así como en solución, los cinco péptidos muestran afinidad diferente para VEGF. Sin embargo, la clasificación de péptidos en solución y en fase gas es invertida.

Este resultado sugiere que en el nivel actual de comprensión del mecanismo de interacciones no covalentes, su energía en fase gas no puede ser tomada como una estimación de la energía de union en solucion, ni cuantitativa ni cualitativamente. Por otro lado, esto muestra el potencial de la espectrometría de masas para proporcionar informacion sobre

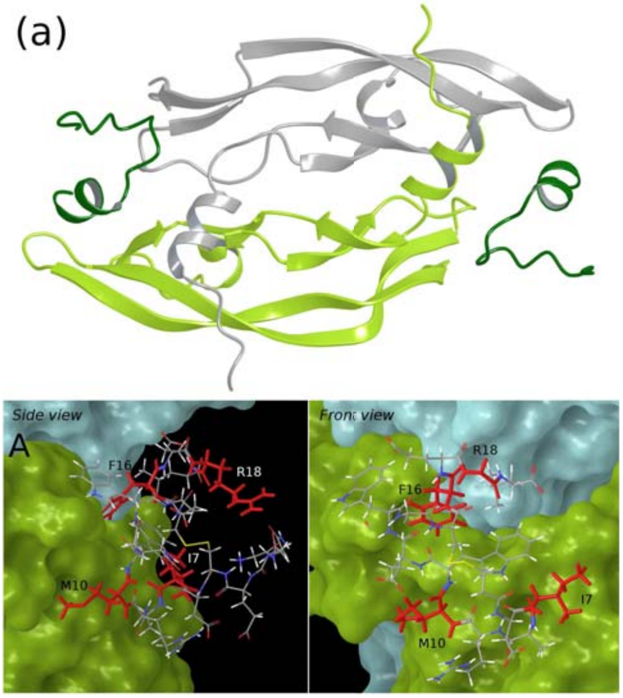


Figura 8.2. (a) Estructura de VEGF con péptido unido **P-wt** obtenida por RMN; (b) primer plano del péptido **P-wt** unido con los residuos que fueron modificados posteriormente destacados en rojo.

Tabla 8.1. lista de péptidos cíclicos usados como ligandos para VEGF. Letras en minúscula y negrita representan D amino ácidos.

Name	***Sequence***
P-wt	GGNECDIARMWEWECFERL
P-7i	GGNECD**i**ARMWEWECFERL
P-18r	GGNECDIARMWEWECFE**r**L
P-16f	GGNECDIARMWEWEC**f**ERL
P-10m	GGNECDIAR**m**WEWECFERL

Tabla 8.2. lista de ligandos peptídicos con sus correspondientes valores de constantes de disociasión y voltajes de colisión.

Name	K_D, *μM*	V_{C50}, *V*
P-wt	*1.02 ± 0.18*	*24.93 ± 0.11*
P-18r	*3.50 ± 0.60*	*27.64 ± 0.20*
P-7i	*252 ± 77*	*25.17 ± 0.14*
P-16f	*313 ± 53*	*28.00 ± 0.18*
P-10m	*1810 ± 147*	*34.31 ± 0.10*

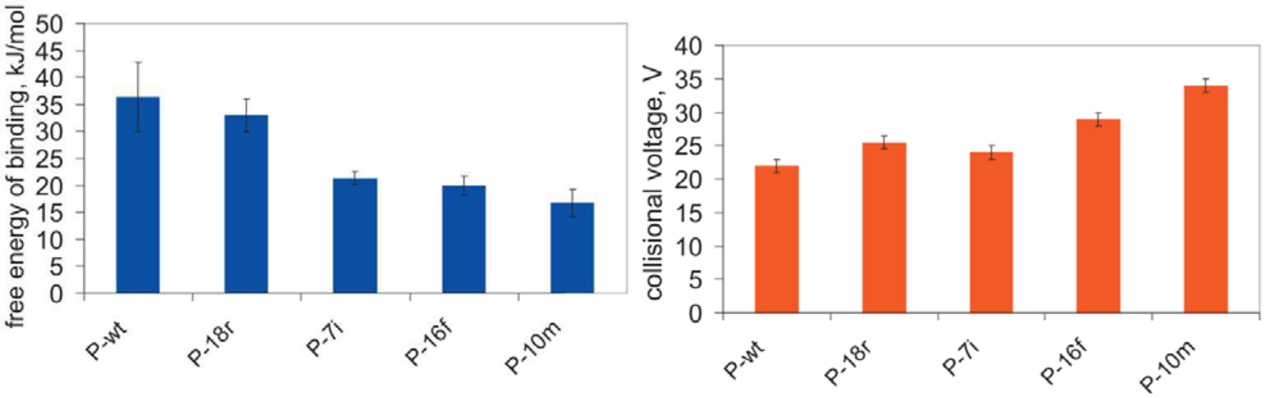
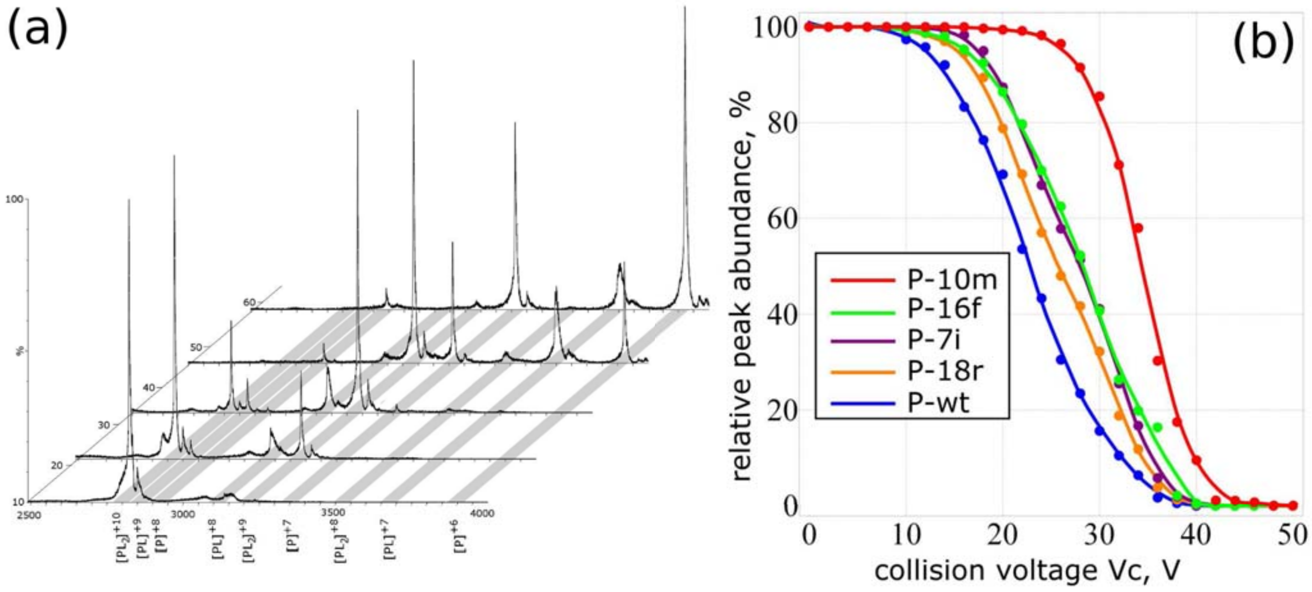


Figura 8.3. arriba: las energías libres de unión de los 5 péptidos respecto a VEGF; abajo: voltajes de colisión necesarios para el 50% de disociación del complejo no-covalente de VEGF con los 5 péptidos en fase gas.



La figura 8.4. (a) Disociación del ión complejo de VEGF PL_2^{+10}. Los diferentes espectros representan diferentes voltajes de colisión, trazados a lo largo del eje de ordenadas; (b) las curvas de disociación de los complejos de VEGF con los 5 péptidos, correspondiente al camino de disociación $PL_2^{+10} \rightarrow PL^{+8} + L^{+2}$.

la naturaleza de las interacciones no covalentes.

Es evidente que en fase gaseosa, así como en solución, los cinco péptidos muestran afinidad diferente para VEGF. Sin embargo, la clasificación de péptidos en solución y en fase gas es invertida.

Este resultado sugiere que en el nivel actual de comprensión del mecanismo de interacciones no covalentes, su energía en fase gas no puede ser tomada como una estimación de la energía de union en solucion, ni cuantitativa ni cualitativamente. Por otro lado, esto muestra el potencial de la espectrometría de masas para proporcionar informacion sobre la naturaleza de las interacciones no covalentes.

8.3.2 Estrategia combinada experimental/computacional para la determinación de la energia de ión en fase gas.

En el artículo 1 mostramos que la espectrometría de masas es potente para proporcionar información adicional sobre el mecanismo de interacciones no covalentes. En el artículo 2 describimos un método que permite al modelaje de la excitación/relajación colisional de iones grandes durante CID. Así mismo, proporciona el medio de calcular la energía interna que los iones adquieren en condiciones arbitrarias experimentales.

Para modelar las colisiones del ión, usamos el método de Monte Carlo. El esquema del algoritmo de Monte Carlo se muestra en la Figura 6A. Se pueden encontrar mas detalles en el aportado “supporting information” del articulo 2.

Los datos experimentales usados en este trabajo fueron adquiridos en el equipo WATERS SYNAPT G1. CID en este instrumento se produce en “la guia de la onda de viaje del ión” (travelling wave ion guide, TWIG). El principio de operación de TWIG se muestra esquemáticamente sobre las Figuras 8.5B y C. El algoritmo para modelar las trayectorias de ión en TWIG esta descrito detalladamente en el aportado “supporting information” del articulo 2. El método fue validado con leu-enkephalin (un neuropeptido de 5 amino ácidos con secuencia YGGFL y peso molecular 556.28 Da). El valor de energía de activación calculado por el método fue de 1.01 ± 0.05 eV, lo cual es cercano al valor publicado (1.14 ± 0.05 eV).

Después de la validación con leu-enkephalin el modelo fue aplicado al conjunto de complejos no covalentes entre VEGF y los 5 péptidos descritos en el artículo 1. Los resultados se muestran en la Figura 8.6. Las curvas finales están representados en la Figura 8.6B, y los valores correspondientes de la energía de activación para los cinco complejos están descritos en la Tabla 8.3.

Tabla 8.3. Lista de péptidos, usados cómo ligandos de VEGF con V_{C50} y valores de las energías de activación calculadas.

Name	***V_{C50}, V***	***E_A, eV***
P-wt	24.93 ± 0.11	8.64 ± 0.63
P-7i	25.17 ± 0.14	8.31 ± 0.61
P-18r	27.64 ± 0.20	9.74 ± 0.71
P-16f	28.00 ± 0.18	9.58 ± 0.70
P-10m	34.31 ± 0.10	13.73 ± 1.00

En general, el trabajo presentado en los artículos 1 y 2 proporciona un entendimiento actual sobre en el estado del arte en el campo de interacciones no covalentes en fase gas. Presentamos un sistema modelo adecuado para el estudio de estos fenómenos que permite llevar a cabo una evaluación preliminar cualitativa de dichos interacciones tanto en solución como fase gas. En el artículo 2 desarrollamos un acercamiento que proporciona la solución a algunas de estas preguntas, cómo, por ejemplo, cuánto energía adquiere el ión a través de colisiones antes de la disociación. Pero hay todavía muchas preguntas por contestar en este contexto antes de que los interacciones no covalentes fase gas sean completamente entendidas. Esto será sujeto de un futuro trabajo.

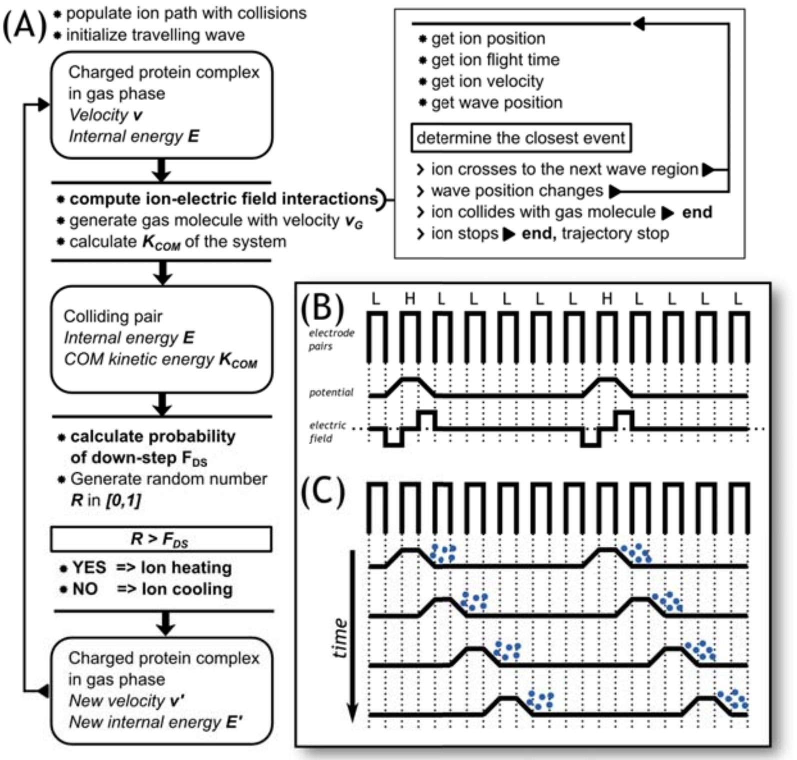


Figura 8.5. (A) Esquema principal del motor de Monte Carlo; (B) representación esquemática de la geometría de parte de la guía de ión TWIG, con forma de potencial y campo; (C) la esquema del transporte del ión en TWIG;

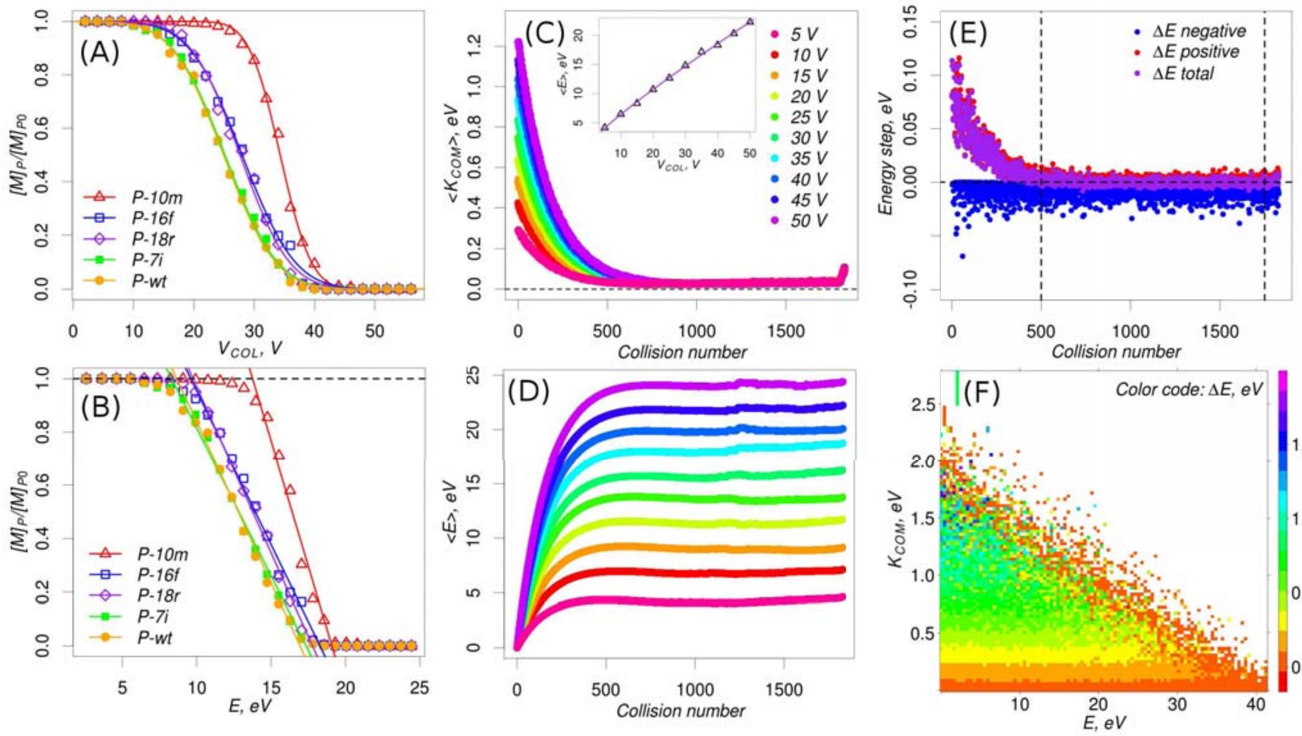


Figura 8.6. Tratamiento de los 5 complejos no covalentes de VEGF con péptidos; (A) curvas de ruptura dependientes del voltaje de colisión experimental (puntos) y curvas ajustadas (líneas); (B) curvas de ruptura dependientes de la energía interna (puntos) y ajuste de sus segmentos lineales (líneas); (C) promedio de la energía cinética en sistema del centro de masas del ión, trazado contra el número de colisiones; (D) promedio de la energía interna del ión, trazado contra el número de colisiones; (E) pasos de energía promedios negativos, positivos y totales trazados contra número de colisiones. (F) gráfico de los pasos de energía, generados durante diez simulación, trazados sobre una rejilla de energías cinéticas y internas.

8.3.3 Uso de espectrometría de masas estructural nativa al estudio de interacciones alostéricas entre chaperonina molecular GroEL y ATP.

(Ranit Gruber y Liat Shimon expresaron las muestras de GroEL que fueron usadas en este trabajo; El Prof. Amnon Horovitz realizó el análisis final de los datos correspondientes a las regulaciones allostericos de GroEL por ATP)

En el artículo 3 presentamos un ejemplo de cómo los rasgos únicos de la espectrometría de masas pueden ayudar a contestar preguntas biológicas relevantes sobre el ejemplo de las interacciones GroEL - ATP. El ciclo de función de GroEL implica el cambio conformacional que es mediado por la unión y la hidrólisis de ATP. Esto sugiere que el estudio de las diferentes formas unidas de GroEL mediante la espectrometría de masas en condiciones dónde la hidrólisis ATP es suprimida, puede proporcionar la información sobre el comportamiento real de GroEL.

Un tampón nuevo compatible con la espectrometría de masas nativa fue descubierto – acetato de etilendiamonio ($C_2H_4(NH_2)OCOCH_3$, EDAA). Este tampón permitió alcanzar una resolución superior a la de los otros dos, sin hidrolisis detectado de ATP.

1 μM GroEL en el tampón de EDAA (200mM EDAA, 1mM $Mg_2(CH_3CO_2)_2$, el pH 7.0) fue titulado con ATP en concentraciones entre 0.5 μM y 100 μM. El fragmento del espectro de GroEL con 5 μM de ATP se muestra en la Figura 8.7A. El pico aumentado de GroEL con 5 μM de ATP de carga +58 y su reconstrucción a partir de gausianos se muestran en las Figuras 8.7B y C respectivamente. Un análisis similar fue realizado para tres estados de carga en cada concentración (Figura 8.7D).

Para ser capaz de correlacionar las poblaciones de los estados de unión diferentes, obtenidos a partir de espectrometría de masas, con electrospray, se tiene que considerar los efectos de unión no específica. El método matemático sugerido en [46] fue corregido y ampliado al caso en el cual la intensidad del pico de proteína libre es cero. Entonces fue usado para eliminar los efectos de unión no específica de los resultados experimentales. El gráfico de saturación fraccionaria Y de GroEL con ATP respecto a la concentración ATP se muestra en la Figura 8.7E. Los datos adquiridos permitieron tratar cada uno de los estados de unión individualmente y así se pudo calcular separadamente los constantes K_i para cada uno de los estados de unión y realizar una esquema de equilibrio entre las conformaciones T (baja afinidad para ATP) y la R (alta afinidad para ATP) (Figura 11F).

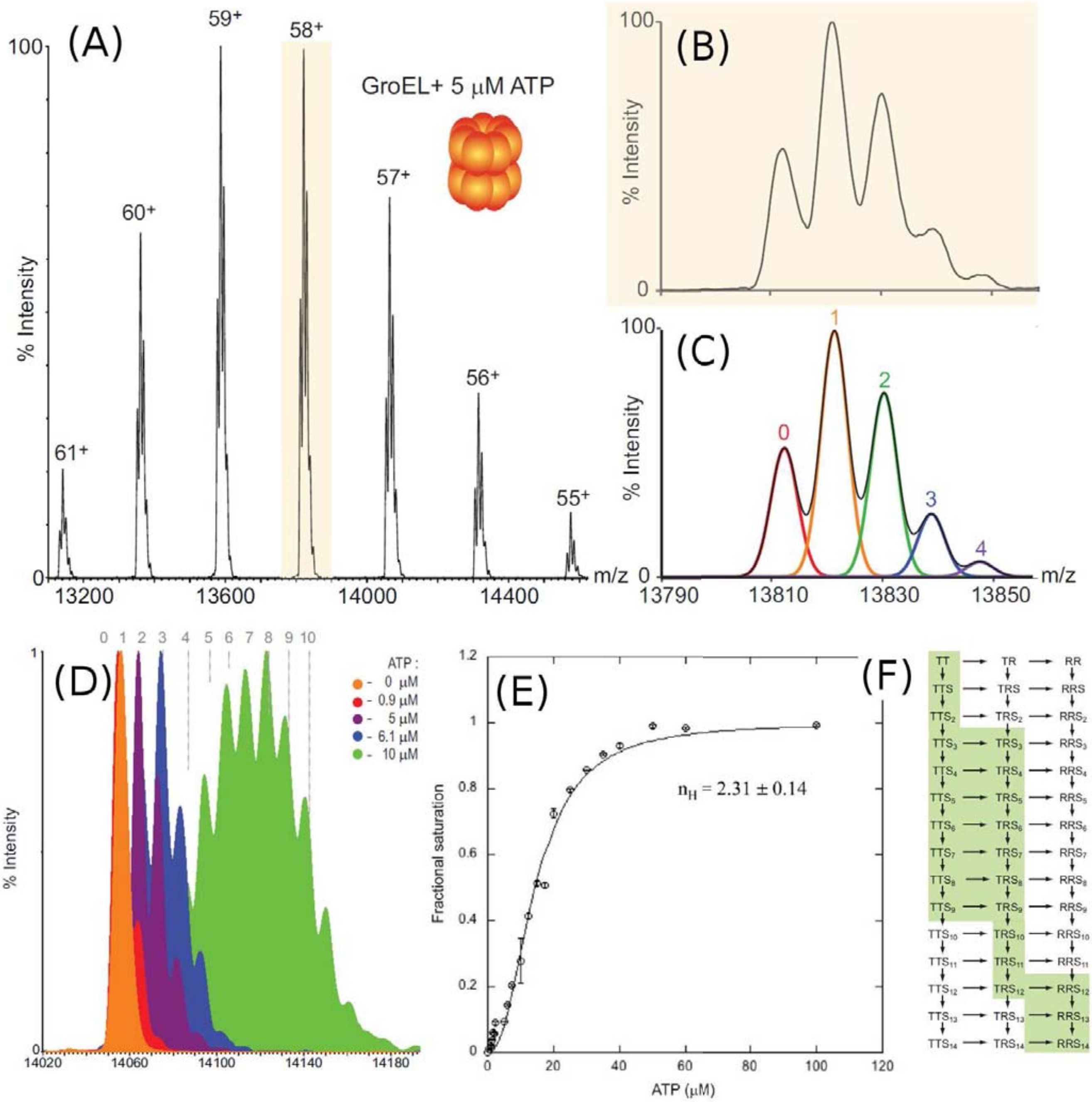


Figura 8.7. (A) Fragmento del espectro de masas ESI-TOF de GroEL + 5 μ M ATP en tampón EDAA; (B) aumento sobre el estado de carga +58; (C) reconstrucción del estado de carga +58 obtenido del mejor ajuste gaussiano; (D) superposición de los espectros del estado de carga +57 de soluciones de GroEL con diferentes concentraciones de ATP; (E) la curva de saturación fraccionaria del complejo unido GroEL-ATP (mirar texto para más detalles); (F) esquema mostrando el equilibrio entre los diferentes conformaciones de los dos anillos de GroEL dependiendo del número de ligandos unidos; las conformaciones más pobladas están destacadas en verde.

8.3.4 Dinámica de la cadena polipeptídica de VEGF estudiada por la combinación del estudio dela relajación por RMN y análisis de modos normales.

(La Dra Margarida Gairi ha detectado los espectros de RMN utilizados para los estudios de relajación magnética)

En este trabajo estudiamos la flexibilidad de la cadena polipeptídica de la proteína VEGF combinando el estudio de la relajación por RMN y el análisis de modos normales. Dos conjuntos de parámetros de relajación de RMN fueron adquiridos. El tiempo de relajación longitudinal T_1, el tiempo de relajación transversal T_2 y el efecto de Overhauser nuclear 1H-15N (NOE) fue obtenido en campos magnéticos estáticos de 14.4 T (600MHz) y 18.1 T (de 800 MHz). Los datos fueron procesados y analizados mediante el programa nmrPipe. Los resultados se muestran en las Figuras 8.8 A, B, C.

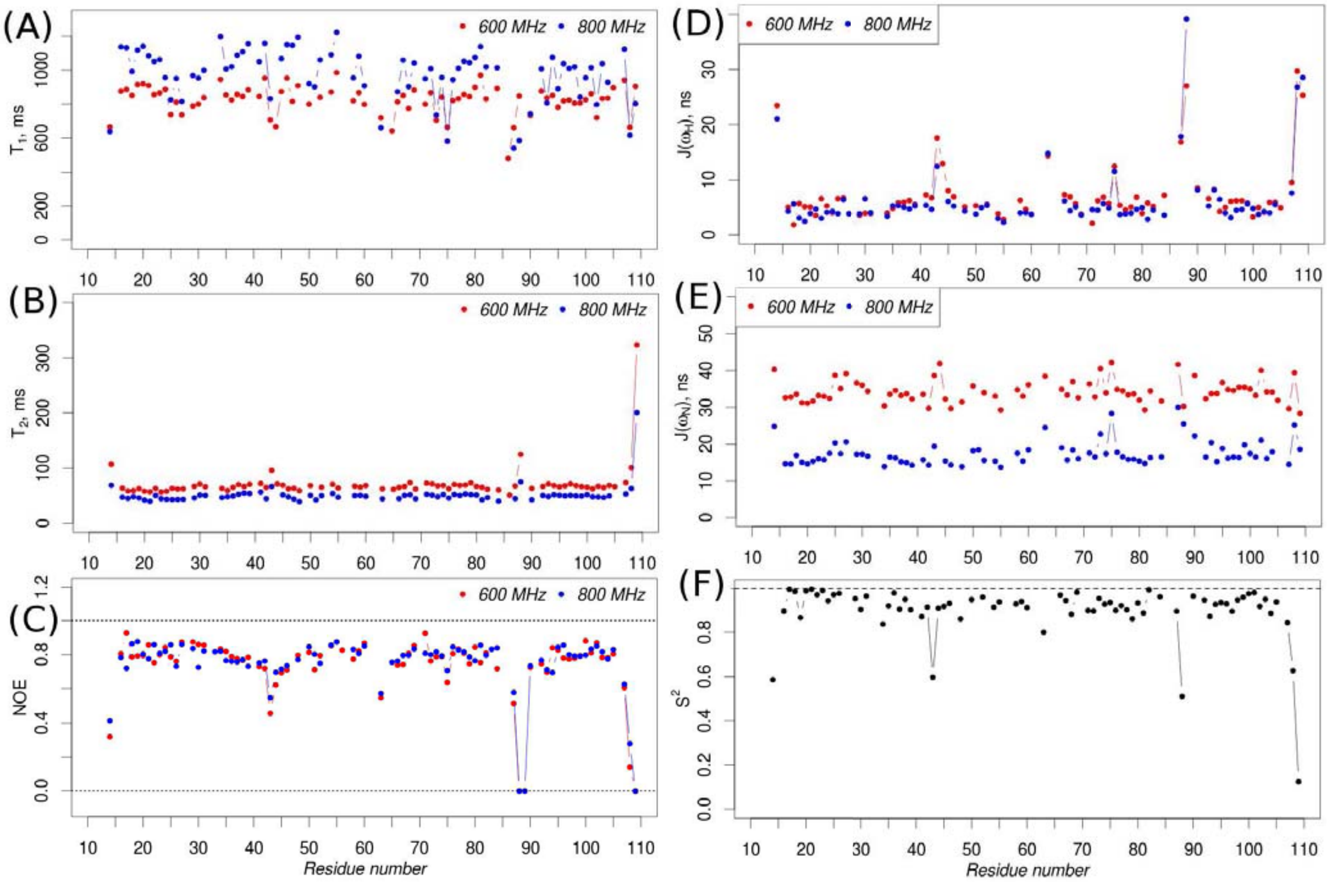


Figura 8.8. (A, B, C) Parámetros de relajación de RMN, adquiridos para VEGF en 45ºC en dos campos estáticos magnéticos: 14.4 T (600 MHz) y 18.1 T (800 MHz), T_1, T_2 y ^{1}H-^{15}N-NOE respectivamente; (D, E) densidades espectrales calculadas a partir de los parámetros de relajación de RMN en las frecuencias Larmour ^{1}H y ^{15}N respectivamente; (F) parámetro de orden generalizado, obtenido del análisis ”model-free” de los datos de relajación adquiridos en 14.1 T.

NOE y T_2 claramente indican que los residuos D63 y Q87-G89, situados sobre el lazo 1 y el lazo 2 están implicados en el movimiento en la escala de tiempo ps-ns. Además de esto, NOE y T_2 también indican la posibilidad de movimiento de ps-ns del residuo I43, localizado en una región inestructurada de la cadena polipeptídica de VEGF. No se encontraron pruebas de cambio químico.

El mapeo de la densidad espectral reducido (reduced spectral density mapping, RSDM) y el analysis “model-free” identificaron los siguientes residuos con elevada flexibilidad: E44 (lazo 1), D63 (región inestructurada), Q87, G88 y H90 (lazo 2), y E73 y N75 (ß-hoja - 5) (Figura 8.8 C,D,F). Cómo esperado, esto confirma conclusiones preliminares hechas a partir de T2 y NOE.

Para complementar los conclusiones experimentales, realizamos el analisis de los modos normales basado en campo de fuerza de VEGF. El RMSD de los carbono-α de la cadena polipeptídica de VEGF fueron calculados para cada uno de los 40 primeros modos. El resultado se muestra en la Figura 8.9.

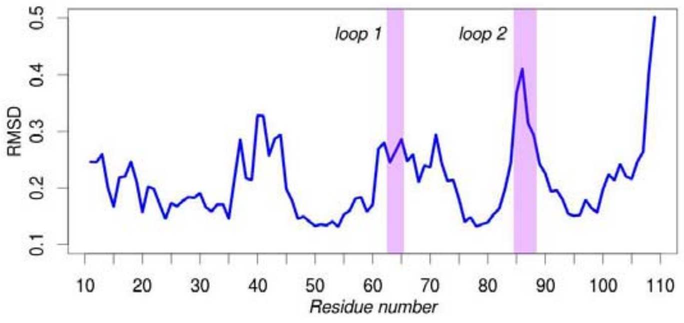


Figura 8.9. promedio de los RMSD derivados del análisis de modos normales. Los dos lazos flexibles son destacados en rosado.

Se puede observar que cuatro áreas flexibles, predichas por análisis de modos normales se superponen en gran parte con los predichos por el estudio de relajación por RMN. Este hecho es mejor ilustrado en la Figura 8.10. Las barras azules y rojas indican las mismas regiones de la cadena polipeptídica de VEGF. Otra observación importante puede ser hecha por la comparación de las áreas flexibles de VEGF con las zonas que están implicadas en interacciones de VEGF-receptor (barras púrpuras en la Figura 8.10). Cómo puede ser visto, la mayor parte de los residuos que está implicados en la unión de VEGF-receptor están situados o se encuentran cerca de las áreas flexibles de cadena polipeptídica VEGF. Esto sugiere que la flexibilidad juega un papel importante en las interacciones no covalentes entre VEGF y sus dos receptores.

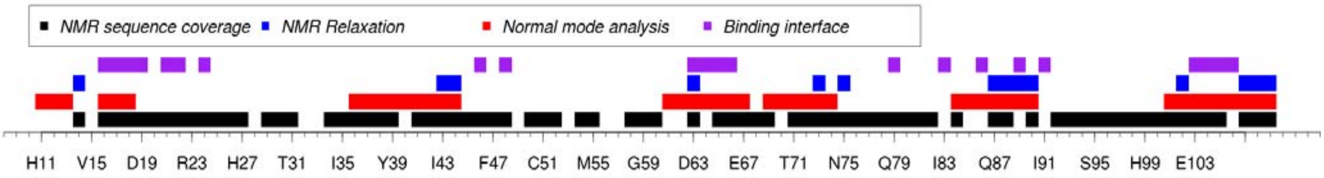


Figura 8.10. Comparación de las regiones flexibles de la cadena polipeptídica de VEGF, predicha por relajación de RMN (azul) y análisis de modos normales (rojo) con residuos, implicados en interacciones de VEGF-receptor (púrpuras). La barra negra indica los residuos para los cuales el conjunto de tres parámetros de relajación estaba disponible en 600 MHz, 800 MHz o ambos.

8.4

Conclusiones

8.4 Conclusiones

A la vista de los resultados presentados, las conclusiones principales de la tesis son:

- La estabilidad en fase gas de los complejos proteina-ligando no se correlaciona, ni cualitativa ni cuantitativamente, con los constantes de afinidad en solución de los mismos complejos, tal como se ha mostrado en el ejemplo de las interacciones no covalentes de VEGF con cinco péptidos cíclicos.
- El análisis de interacciones no covalentes mediante espectrometría de masas puede proporcionar información cualitativa valiosa sobre el papel del disolvente. Sin embargo, para obtener información cuantitativa, se necesita un método robusto para computar los estabilidades de los complejos no covalentes en fase gas.
- El método combinado experimental/computacional, descrito en la presente tesis, permite modelar la activación colisional de iones de moléculas o complejos de elevado peso molecular en procesos de CID. El método es universal y puede ser aplicado a diferentes sistemas sin límite teórico superior para el peso molecular. Sin embargo, está diseñado para funcionar mejor con iones de gran tamaño.
- La aproximación desarrollada también se puede aplicar a los distintos tipos de cámaras de colisión usadas en los espectrómetros de masas comerciales. En concreto, en este trabajo se ha aplicada a la simulación de trayectorias de iones en TWIG, del equipo SYNAPT G1 de Waters.
- Se ha aplicado la espectrometría de masas nativa al estudio de interacciones allostericas entre GroEL y ATP. Se ha modelado el equilibrio entre las dos conformaciones de GroEL fue variando el numero de moléculas ATP unidas a la proteína. Los resultados obtenidos están a favor del modelo de cooperación MWC.
- A lo largo de este estudio se ha descubierto un nuevo tampón compatible con espectrometría de masas nativa (acetato de etilendiamonio). El uso de este nuevo medio ha permitido mejorar las resoluciones alcanzadas con los tampones que se usan tradicionalmente.
- A lo largo del estudio de la interacción GroEL y ATP se ha corregido y ampliado la estrategia matemática, descrita previamente en la literatura, para la supresión de de los efectos de unión no especifica.
- Se ha estudiada la dinámica del esqueleto polipeptídico de VEGF mediante el uso combinado de estudios de relajación magnética y análisis de modos normales (NMA). Se ha comprobado que NMA puede proporcionar información relevante sobre la flexibilidad de proteína y complementar el estudio de relajación por RMN.
- Utilizando la metodología descrita en el punto anterior se han identificado cuatro zonas flexibles en la superficie de VEGF. Se ha observado que dichas zonas flexibles coinciden en gran parte con la zona de interaccion entre VEGF y sus receptores.

Referencias:

[1] B. Pan, B. Li, S. J. Russell, J. Y. K. Tom, A. G. Cochran, and W. J. Fairbrother, “Solution structure of a phage-derived peptide antagonist in complex with vascular endothelial growth factor.,” *Journal of molecular biology*, vol. 316, no. 3, pp. 769–87, Feb. 2002.

[2] S. Ilchenko and R. J. Cotter, “Collision Energetics in a Tandem Time-of-Flight (TOF/TOF) Mass Spectrometer with a Curved-Field Reflectron.,” *International journal of mass spectrometry*, vol. 265, no. 2–3, pp. 372–381, Sep. 2007.

[3] M. Baker, “Proteomics: The interaction map,” *Nature*, pp. 2–5, 2012.

[4] S. Gordo and E. Giralt, “Knitting and untying the protein network: modulation of protein ensembles as a therapeutic strategy.,” *Protein science : a publication of the Protein Society*, vol. 18, no. 3, pp. 481–93, Mar. 2009.

[5] P. L. Kastritis, I. H. Moal, H. Hwang, Z. Weng, P. A. Bates, A. M. J. J. Bonvin, and J. Janin, “A structure-based benchmark for protein-protein binding affinity.,” *Protein science : a publication of the Protein Society*, vol. 20, no. 3, pp. 482–91, Mar. 2011.

[6] G. Klebe and H. J. Böhm, “Energetic and entropic factors determining binding affinity in protein-ligand complexes.,” *Journal of receptor and signal transduction research*, vol. 17, no. 1–3, pp. 459–73.

[7] M. K. Gilson and B. Honig, “Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis.,” *Proteins*, vol. 4, no. 1, pp. 7–18, Jan. 1988.

[8] K. Henzler-Wildman and D. Kern, “Dynamic personalities of proteins.,” *Nature*, vol. 450, no. 7172, pp. 964–72, Dec. 2007.

[9] A. Newman, “Product Review: The Precise World of Isotope Ratio Mass Spectrometry,” *Analytical Chemistry*, vol. 68, no. 11, p. 373A–377A, Jun. 1996.

[10] C. Uetrecht, C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck, “High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 27, pp. 9216–20, Jul. 2008.

[11] T. Kind and O. Fiehn, “Advances in structure elucidation of small molecules using mass spectrometry.,” *Bioanalytical reviews*, vol. 2, no. 1–4, pp. 23–60, Dec. 2010.

[12] C. Lifshitz and J. Laskin, *Principles of mass spectrometry applied to biomolecules*. John Wiley & Sons, Inc., 2006.

[13] M. Sharon, “How far can we go with structural mass spectrometry of protein complexes?,” *Journal of the American Society for Mass Spectrometry*, vol. 21, no. 4, pp. 487–500, Apr. 2010.

[14] K. Vékey, “Internal Energy Effects in Mass Spectrometry,” *Journal of Mass Spectrometry*, vol. 31, no. 5, pp. 445–463, May 1996.

[15] A. J. R. Heck, “Native mass spectrometry: a bridge between interactomics and structural biology.,” *Nature methods*, vol. 5, no. 11, pp. 927–33, Nov. 2008.

[16] J. Lengqvist, A. Mata de Urquiza, T. Perlmann, J. Sjövall, and W. J. Griffiths, “Specificity of receptor-ligand interactions and their effect on dimerisation as observed by electrospray mass spectrometry: bile acids form stable adducts to the RXRalpha.,” *Journal of mass spectrometry : JMS*, vol. 40, no. 11, pp. 1448–61, Nov. 2005.

[17] T. L. Pukala, B. T. Ruotolo, M. Zhou, A. Politis, R. Stefanescu, J. A. Leary, and C. V Robinson, “Subunit architecture of multiprotein assemblies determined using restraints from gas-phase measurements.,” *Structure (London, England : 1993)*, vol. 17, no. 9, pp. 1235–43, Sep. 2009.

[18] H. J. Maple, R. A. Garlish, L. Rigau-Roca, J. Porter, I. Whitcombe, C. E. Prosser, J. Kennedy, A. J. Henry, R. J. Taylor, M. P. Crump, and J. Crosby, “Automated protein-ligand interaction screening by mass spectrometry.,” *Journal of medicinal chemistry*, vol. 55, no. 2, pp. 837–51, Jan. 2012.

[19] J. M. Daniel, S. D. Friess, S. Rajagopalan, S. Wendt, and R. Zenobi, “Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry,” *International Journal of Mass Spectrometry*, vol. 216, no. 1, pp. 1–27, Apr. 2002.

[20] P. M. Mayer and E. Martineau, “Gas-phase binding energies for non-covalent Aβ-40 peptide/small molecule complexes from CID mass spectrometry and RRKM theory.,” *Physical chemistry chemical physics : PCCP*, vol. 13, no. 11, pp. 5178–86, Mar. 2011.

[21] V. J. Nesatyy, “Gas-phase binding of non-covalent protein complexes between bovine pancreatic trypsin inhibitor and its target enzymes studied by electrospray ionization tandem mass spectrometry.,” *Journal of mass spectrometry : JMS*, vol. 36, no. 8, pp. 950–9, Aug. 2001.

[22] M. Schäfer, C. Schmuck, M. Heil, H. J. Cooper, C. L. Hendrickson, M. J. Chalmers, and A. G. Marshall, “Determination of the activation energy for unimolecular dissociation of a non-covalent gas-phase peptide: substrate complex by infrared multiphoton dissociation fourier transform ion cyclotron resonance mass spectrometry.,” *Journal of the American Society for Mass Spectrometry*, vol. 14, no. 11, pp. 1282–9, Nov. 2003.

[23] W. Wang, E. N. Kitova, and J. S. Klassen, “Bioactive recognition sites may not be energetically preferred in protein-carbohydrate complexes in the gas phase.,” *Journal of the American Chemical Society*, vol. 125, no. 45, pp. 13630–1, Nov. 2003.

[24] K. Breuker, “The study of protein–ligand interactions by mass spectrometry—a personal view,” *International Journal of Mass Spectrometry*, vol. 239, no. 1, pp. 33–41, Dec. 2004.

[25] A. Mittermaier and L. E. Kay, “New tools provide new insights in NMR studies of protein dynamics.,” *Science (New York, N.Y.)*, vol. 312, no. 5771, pp. 224–8, Apr. 2006.

[26] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, and D. Kern, “Intrinsic motions along an enzymatic reaction trajectory.,” *Nature*, vol. 450, no. 7171, pp. 838–44, Dec. 2007.

[27] O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot, “Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution.,” *Science (New York, N.Y.)*, vol. 320, no. 5882, pp. 1471–5, Jun. 2008.

[28] L.-W. Yang, a J. Rader, X. Liu, C. J. Jursa, S. C. Chen, H. a Karimi, and I. Bahar, “oGNM: online computation of structural dynamics using the Gaussian Network Model.,” *Nucleic acids research*, vol. 34, no. Web Server issue, pp. W24–31, Jul. 2006.

[29] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar, “PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique,” *Journal of Chemical Theory and Computation*, vol. 1, no. 6, pp. 1304–1311, Nov. 2005.

[30] D. Motiejunas, R. Gabdoulline, T. Wang, A. Feldman-Salit, T. Johann, P. J. Winn, and R. C. Wade, “Protein-protein docking by simulating the process of association subject to biochemical constraints.,” *Proteins*, vol. 71, no. 4, pp. 1955–69, Jun. 2008.

[31] A. G. Palmer, “NMR characterization of the dynamics of biomacromolecules.,” *Chemical reviews*, vol. 104, no. 8, pp. 3623–40, Aug. 2004.

[32] R. Ishima and K. Nagayama, “Protein Backbone Dynamics Revealed by Quasi Spectral

Density Function Analysis of Amide N-15 Nuclei," *Biochemistry*, vol. 34, no. 10, pp. 3162–3171, Mar. 1995.

[33] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results," *Journal of the American Chemical Society*, vol. 2, no. 1, 1982.

[34] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity," *Journal of the American Chemical Society*, vol. 104, no. 17, pp. 4546–4559, Aug. 1982.

[35] S. Iyer and K. R. Acharya, "Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines.," *The FEBS journal*, vol. 278, no. 22, pp. 4304–22, Nov. 2011.

[36] Y. a Muller, H. W. Christinger, B. a Keyt, and a M. de Vos, "The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 A resolution: multiple copy flexibility and receptor binding.," *Structure (London, England : 1993)*, vol. 5, no. 10, pp. 1325–38, Oct. 1997.

[37] Y. a Muller, B. Li, H. W. Christinger, J. a Wells, B. C. Cunningham, and a M. de Vos, "Vascular endothelial growth factor: crystal structure and functional mapping of the kinase domain receptor binding site.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 14, pp. 7192–7, Jul. 1997.

[38] B. A. C. Horta, J. J. V. Cirino, and R. B. de Alencastro, "On the structure, interactions, and dynamics of bound VEGF.," *Journal of molecular graphics & modelling*, vol. 26, no. 7, pp. 1091–103, Apr. 2008.

[39] N. Ferrara, "The Biology of Vascular Endothelial Growth Factor," *Endocrine Reviews*, vol. 18, no. 1, pp. 4–25, Feb. 1997.

[40] N. Lu, Y. Gao, Y. Ling, Y. Chen, Y. Yang, H.-Y. Gu, Q. Qi, W. Liu, X.-T. Wang, Q.-D. You, and Q.-L. Guo, "Wogonin suppresses tumor growth in vivo and VEGF-induced angiogenesis through inhibiting tyrosine phosphorylation of VEGFR2.," *Life sciences*, vol. 82, no. 17–18, pp. 956–63, Apr. 2008.

[41] W. D. Brazelle, W. Shi, and D. W. Siemann, "VEGF-associated tyrosine kinase inhibition increases the tumor response to single and fractionated dose radiotherapy.," *International journal of radiation oncology, biology, physics*, vol. 65, no. 3, pp. 836–41, Jul. 2006.

[42] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Vascular endothelial growth factor (VEGF) inhibition--a critical review.," *Anti-cancer agents in medicinal chemistry*, vol. 7, no. 2, pp. 223–45, Mar. 2007.

[43] K. A. Krishna, G. V. Rao, and K. R. S. S. Rao, "Chaperonin GroEL: structure and reaction cycle.," *Current protein & peptide science*, vol. 8, no. 5, pp. 418–25, Oct. 2007.

[44] P. B. Sigler, Z. Xu, H. S. Rye, S. G. Burston, W. A. Fenton, and A. L. Horwich, "Structure and function in GroEL-mediated protein folding.," *Annual review of biochemistry*, vol. 67, pp. 581–608, Jan. 1998.

[45] H. Taguchi, "Chaperonin GroEL meets the substrate protein as a 'load' of the rings.," *Journal of biochemistry*, vol. 137, no. 5, pp. 543–9, May 2005.

[46] L. Shimon, M. Sharon, and A. Horovitz, "A method for removing effects of nonspecific binding on the distribution of binding stoichiometries: application to mass spectroscopy data.," *Biophysical journal*, vol. 99, no. 5, pp. 1645–9, Sep. 2010.